

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:19:10 ; Search time 9.98505 Seconds
(without alignments)
1946.896 Million cell updates/sec

Title: US-09-988-117-3
Perfect score: 5301
Sequence: 1 MAOETSSSRLLGPGCEPAAE.....SISPEGUSPPRRKDDPTMP 1002

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	5301 100.0	1002 9 US-09-988-117-3	Sequence 3, Appl1
2	5301 100.0	1002 10 US-09-812-471-3	Sequence 3, Appl1
3	5301 100.0	1002 10 US-09-812-633-3	Sequence 3, Appl1
4	4677.5 88.2	1003 9 US-09-988-117-1	Sequence 1, Appl1
5	4677.5 88.2	1005 10 US-09-812-471-1	Sequence 1, Appl1
6	4677.5 88.2	1005 10 US-09-812-633-1	Sequence 1, Appl1
7	482.5 9.1	330 10 US-09-864-761-42953	Sequence 42953, A
8	438.5 8.3	336 10 US-09-864-761-43157	Sequence 43157, A
9	297.5 5.6	515 10 US-09-864-761-37152	Sequence 37152, A
10	270.5 5.1	577 10 US-09-764-864-883	Sequence 883, App
11	264 5.0	610 9 US-09-989-920-212	Sequence 212, App
12	249.5 4.7	406 10 US-09-764-864-976	Sequence 976, App
13	244.5 4.6	375 10 US-09-864-761-36336	Sequence 36336, A
14	244.5 4.6	513 10 US-09-764-864-1061	Sequence 1061, Ap
15	242.5 4.6	439 10 US-09-864-761-33476	Sequence 33476, A
16	242 4.6	580 10 US-09-925-301-1357	Sequence 1357, Ap
17	242 4.6	2005 10 US-09-735-367B-3	Sequence 3, Appl1
18	242 4.6	2063 10 US-09-735-367B-2	Sequence 2, Appl1
19	241.5 4.6	338 10 US-09-764-864-948	Sequence 948, App

20	239.5	4.5	1325	10	US-09-864-761-35612	Sequence 35612, A
21	238	4.5	342	10	US-09-864-761-36708	Sequence 36708, A
22	237	4.5	551	10	US-09-864-761-37065	Sequence 37065, A
23	236.5	4.5	2783	10	US-09-816-669A-14	Sequence 14, Appl1
24	235.5	4.4	536	10	US-09-864-761-36148	Sequence 36148, A
25	234	4.4	311	10	US-09-864-761-34333	Sequence 34333, A
26	234	4.4	340	10	US-09-864-761-35894	Sequence 35894, A
27	233.5	4.4	983	10	US-09-864-761-35800	Sequence 38000, A
28	232	4.4	361	10	US-09-764-864-893	Sequence 893, App
29	231	4.4	708	10	US-09-764-864-954	Sequence 954, App
30	229	4.3	428	9	US-09-734-329-2	Sequence 2, Appl1
31	227.5	4.3	525	10	US-09-864-761-33653	Sequence 33653, A
32	226	4.3	378	10	US-09-764-864-956	Sequence 956, App
33	226	4.3	492	10	US-09-864-761-42897	Sequence 42897, A
34	226	4.3	516	9	US-09-974-298-118	Sequence 118, App
35	225.5	4.3	383	10	US-09-764-864-1001	Sequence 1001, Ap
36	224.5	4.2	252	10	US-09-816-669A-12	Sequence 12, Appl
37	224.5	4.2	543	9	US-09-736-457-337	Sequence 337, App
38	224.5	4.2	543	9	US-09-902-941-337	Sequence 337, App
39	224.5	4.2	543	9	US-09-849-626-337	Sequence 337, App
40	222	4.2	449	10	US-09-929-315-4	Sequence 4, Appl1
41	222	4.2	449	10	US-09-929-315-4	Sequence 4, Appl1
42	221	4.2	558	10	US-09-864-761-47724	Sequence 47724, A
43	218.5	4.1	284	10	US-09-764-864-932	Sequence 932, App
44	218	4.1	468	10	US-09-864-761-37991	Sequence 37991, A
45	216.5	4.1	510	10	US-09-866-562-56	Sequence 56, Appl
46	216.5	4.1	547	10	US-09-764-864-1065	Sequence 1065, Ap
47	216	4.1	1464	9	US-09-842-777-10	Sequence 10, Appl
48	215.5	4.1	449	10	US-09-929-315-6	Sequence 6, Appl1
49	215.5	4.1	449	10	US-09-929-315-6	Sequence 6, Appl1
50	215	4.1	269	10	US-09-764-864-887	Sequence 887, App
51	214	4.0	333	10	US-09-764-864-947	Sequence 947, App
52	214	4.0	333	10	US-09-764-864-1095	Sequence 1095, Ap
53	214	4.0	448	10	US-09-764-864-909	Sequence 909, App
54	212	4.0	600	10	US-09-764-864-957	Sequence 957, App
55	210	4.0	316	10	US-09-764-864-1036	Sequence 1036, Ap
56	209	3.9	483	10	US-09-764-864-940	Sequence 940, App
57	209	3.9	516	10	US-09-083-290-4	Sequence 4, Appl1
58	208	3.9	432	10	US-09-755-830-39	Sequence 39, Appl
59	208	3.9	506	10	US-09-083-290-2	Sequence 2, Appl1
60	208	3.9	537	9	US-10-037-667-1	Sequence 1, Appl1
61	208	3.9	1125	9	US-09-974-288-114	Sequence 114, App
62	207	3.9	281	10	US-09-764-864-1337	Sequence 1337, Ap
63	206	3.9	614	10	US-09-864-761-43107	Sequence 43107, A
64	204.5	3.9	488	10	US-09-840-787-17	Sequence 17, Appl
65	203.5	3.8	359	10	US-09-764-864-996	Sequence 996, App
66	203.5	3.8	585	10	US-09-764-864-1063	Sequence 1063, Ap
67	203	3.8	261	10	US-09-764-864-961	Sequence 961, App
68	203	3.8	461	10	US-09-755-850-38	Sequence 38, Appl
69	201.5	3.8	515	9	US-10-037-667-4	Sequence 4, Appl1
70	200	3.8	202	10	US-09-764-864-1337	Sequence 1337, Ap
71	199	3.8	241	10	US-09-764-864-935	Sequence 935, App
72	197.5	3.7	196	9	US-09-942-087A-30	Sequence 30, Appl
73	197.5	3.7	196	10	US-09-779-233-18	Sequence 18, Appl
74	197.5	3.7	196	10	US-09-844-508-29	Sequence 29, Appl
75	197.5	3.7	413	10	US-09-764-864-995	Sequence 995, App
76	197	3.7	826	10	US-09-894-998-47	Sequence 47, Appl
77	195.5	3.7	571	10	US-09-764-864-1028	Sequence 1028, Ap
78	195	3.7	711	10	US-09-976-165-10	Sequence 10, Appl
79	195	3.7	711	10	US-09-828-648-2	Sequence 2, Appl1
80	195	3.7	731	9	US-10-086-464-8	Sequence 8, Appl1
81	194.5	3.7	249	10	US-09-925-300-1179	Sequence 1179, Ap
82	194	3.7	273	10	US-09-764-864-1421	Sequence 1421, Ap
83	193.5	3.7	532	9	US-10-037-667-2	Sequence 2, Appl1
84	193	3.6	277	10	US-09-864-761-36043	Sequence 36043, A
85	193	3.6	361	10	US-09-955-518-21	Sequence 21, Appl
86	193	3.6	481	10	US-09-864-761-37651	Sequence 37651, A
87	192.5	3.6	498	9	US-10-037-667-5	Sequence 5, Appl1
88	192	3.6	488	9	US-09-964-899-35	Sequence 35, Appl
89	192	3.6	723	12	US-10-044-205A-32	Sequence 32, Appl
90	192	3.6	1907	9	US-09-832-292-39	Sequence 39, Appl
91	191.5	3.6	1907	10	US-09-764-864-991	Sequence 991, App
92	191.5	3.6	796	12	US-10-044-205A-31	Sequence 31, Appl

93	190.5	3.6	1367	10	US-09-801-368-108	Sequence 108, App
94	190.5	3.6	4019	9	US-09-854-133-425	Sequence 425, App
95	190.5	3.6	4019	10	US-09-738-973-425	Sequence 425, App
96	190	3.6	322	10	US-09-764-664-014	Sequence 911, App
97	189.5	3.6	605	10	US-09-864-676-1-65132	Sequence 36132, A
98	189.5	3.6	605	10	US-09-864-761-37943	Sequence 37943, A
99	189.5	3.6	944	10	US-09-864-238-2	Sequence 2, Appl1
100	189	3.6	1744	9	US-10-108-605-25	Sequence 25, Appl1

ALIGNMENTS

```

RESULT 1
US-09-988-117-3
Sequence 3, Application US/09988117
Patent No. US20020156039A1
GENERAL INFORMATION:
APPLICANT: Benjamin, Thomas L.
APPLICANT: Li, Dawei
APPLICANT: Mok, Samuel C.
APPLICANT: Cramer, Daniel W.
APPLICANT: Ma, Yupo
TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
FILE REFERENCE: 00742/066002
CURRENT APPLICATION NUMBER: US/09/988,117
CURRENT FILING DATE: 2001-11-16
PRIOR APPLICATION NUMBER: US 09/812,633
PRIOR FILING DATE: 2001-03-19
PRIOR APPLICATION NUMBER: US 60/216,723
PRIOR FILING DATE: 2000-07-07
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 1002
TYPE: PRT
ORGANISM: Mus musculus
US-09-988-117-3

```

Query Match	100.0%	Score 5301;	DB 9;	Length 1002;
Best Local Similarity	100.0%	Pred. No. 4.6e-247;		
Matches 1002; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	MAQETGSSSSRLGGPCCEPAERGGDASEEHNPOVCAKCCAQOFSPEETFLAHONSCCTDPV	60
Dp	1	MAQETGSSSSRLGGPCCEPAERGGDASEEHNPOVCAKCCAQOFSPEETFLAHONSCCTDPV	60
QY	61	MYIIGQENPSSNSAASPAREEGHSRSQVMDTEHSPDPSSGGPPDPWTGPERRGESS	120
Dp	61	MYIIGQENPSSNSAASPAREEGHSRSQVMDTEHSPDPSSGGPPDPWTGPERRGESS	120
QY	121	GOFLVAATGTAAGGGGGLILASPKLATPLPEESTPADPPPPPPPPGVGSHLNIPLI	180
Dp	121	GOFLVAATGTAAGGGGGLILASPKLATPLPEESTPADPPPPPPPPGVGSHLNIPLI	180
QY	181	LEELRVLOORLIHOMQTEQICRQVLLLSIGQIYGAPASBELPCTGAASSTKPLILPLF	240
Dp	181	LEELRVLOORLIHOMQTEQICRQVLLLSIGQIYGAPASBELPCTGAASSTKPLILPLF	240
QY	241	SPIRPAQGTGTAASSSSSSSGCAEPKQKQAFNLIYHPIGSGQHPFVGGVGRSHKPTPAPS	300
Dp	241	SPIRPAQGTGTAASSSSSSSGCAEPKQKQAFNLIYHPIGSGQHPFVGGVGRSHKPTPAPS	300
QY	301	PALPGSTQDILIASPHILAFPGTTGILAAOCGAARGLLEAASPGILKPKVGSGBLGGEVI	360
Dp	301	PALPGSTQDILIASPHILAFPGTTGILAAOCGAARGLLEAASPGILKPKVGSGBLGGEVI	360
QY	361	SSLEKPGGRHKCRFCAKVFGSDSALQIHLRSHTGERPKYCNVCGNRFTTTRGNLKVYFHRH	420
Dp	361	SSLEKPGGRHKCRFCAKVFGSDSALQIHLRSHTGERPKYCNVCGNRFTTTRGNLKVYFHRH	420
QY	421	REKYPHQMNHPRVFEHLDIYVITSSGLPIGMSVPRKEAEAEAGTPGGGVERKPLVAASTTA	480

Db	421	REKDYVQMNPRHPRVPHLDYVITSSGLPYGMSVPRPEKAEEREGYGGVERKPRVLAUSTA	480
Qy	481	LSAIESLTLTLSTGTSTAVAPGLPTFENKFVLMKAVEPKSKADENTPPGSGSAIAGVADSG	540
Db	481	LSAIESLTLTLSTGTSTAVAPGLPTFENKFVLMKAVEPKSKADENTPPGSGSAIAGVADSG	540
Qy	541	SATRMQLSKLVTSLSMALLTNMLKSTGSPRPFVLEPRVLGASPSSTSKLOOLVEKIDRGG	600
Db	541	SATRMQLSKLVTSLSMALLTNMLKSTGSPRPFVLEPRVLGASPSSTSKLOOLVEKIDRGG	600
Qy	601	AVAAVASTASGAPPTSAPAPSSSASGECNOCTICLRVLSCPRALRHYGONGGERPKKVC	660
Db	601	AVAAVASTASGAPPTSAPAPSSSASGECNOCTICLRVLSCPRALRHYGONGGERPKKVC	660
Qy	661	GRASTRGNLNAHFVGHKTSPPARAONSCPICOKKFTNAVTILOOHVNMHLGOLPNGGSA	720
Db	661	GRASTRGNLNAHFVGHKTSPPARAONSCPICOKKFTNAVTILOOHVNMHLGOLPNGGSA	720
Qy	721	LSSEGGAQAQENSSEGSTASGPGSPFOQSQSPSEEMSEEEDEDEEEDVTDSDSLAG	780
Db	721	LSSEGGAQAQENSSEGSTASGPGSPFOQSQSPSEEMSEEEDEDEEEDVTDSDSLAG	780
Qy	781	RGSSSGGEKALISVRKDSSEVSGAEEEVATSVAAPTYVKEMDSNEKADQNTLPRPPRPN	840
Db	781	RGSSSGGEKALISVRKDSSEVSGAEEEVATSVAAPTYVKEMDSNEKADQNTLPRPPRPN	840
Qy	841	LDHPQPMQEGTSDVSGAAMEEKALBESSPMALTOEGESTSPYLVELNLNPEAMKXPG	900
Db	841	LDHPQPMQEGTSDVSGAAMEEKALBESSPMALTOEGESTSPYLVELNLNPEAMKXPG	900
Qy	901	ESSGRKACEVCGSGSPPTOTALIEHOKTHPKDGLFTVCFCROGFLDRATLKKHMLAHNQ	960
Db	901	ESSGRKACEVCGSGSPPTOTALIEHOKTHPKDGLFTVCFCROGFLDRATLKKHMLAHNQ	960
Qy	961	VPPFAPHGPNATILSLVPGSSSTLPSPLSPPRRKDDPTMP 1002	
Db	961	VPPFAPHGPNATILSLVPGSSSTLPSPLSPPRRKDDPTMP 1002	
Qy	961	VPPFAPHGPNATILSLVPGSSSTLPSPLSPPRRKDDPTMP 1002	
Db	961	VPPFAPHGPNATILSLVPGSSSTLPSPLSPPRRKDDPTMP 1002	

RESULT 2
US-09-812-471-3
Sequence 3, Application US/09812471

```

? Patent No. US20020018765A1
?
? GENERAL INFORMATION:
? APPLICANT: Benjamin, Thomas L.
? TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
? TITLE OF INVENTION: Using Mutant Viruses
? FILE REFERENCE: 0074/062002
? CURRENT APPLICATION NUMBER: US/09/812,471
? CURRENT FILING DATE: 2001-03-19
? PRIOR APPLICATION NUMBER: US 60/216,723
? PRIOR FILING DATE: 2000-07-07
? NUMBER OF SEQ ID NOS: 21
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO: 3
? LENGTH: 1002
? TYPE: PRT
? ORGANISM: Mus musculus
? US-09-812-471-3

```

Query Match	100.0%	Score 5301;	DB 10;	Length 1002;
Best Local Similarity	100.0%	Pred. NO.4.6e-247;		
Matches 1002; Conservative	0;	Mismatches	0;	Gaps 0;

QY	1	MAOEGSSSRLGCPGCEPAERGGDASEEHHPDYCAKCAAFSDTEFLAHONSCCTDPV	60
QY	1	MAOEGSSSRLGCPGCEPAERGGDASEEHHPDYCAKCAAFSDTEFLAHONSCCTDPV	60
Db	1	MAOEGSSSRLGCPGCEPAERGGDASEEHHPDYCAKCAAFSDTEFLAHONSCCTDPV	60
QY	61	MYLGGQENPNSSSASSAPRPGSHSRQVWDTEHNSNPDDGSSCPPTPTWPPERGESS	120
Db	61	MYLGGQENPNSSSASSAPRPGSHSRQVWDTEHNSNPDDGSSCPPTPTWPPERGESS	120
QY	61	MYLGGQENPNSSSASSAPRPGSHSRQVWDTEHNSNPDDGSSCPPTPTWPPERGESS	120
Db	61	MYLGGQENPNSSSASSAPRPGSHSRQVWDTEHNSNPDDGSSCPPTPTWPPERGESS	120

```
OY 121 GQFLVAAATGTAAGGGGGLILASPKLGATPLPEESTPAPPPPPPPPPPPGVGSHLNIPLI 180
    |||||||
Db 121 GQFLVAAATGTAAGGGGGLILASPKLGATPLPEESTPAPPPPPPPPPPPGVGSHLNIPLI 180
OY 181 LEEELRVLOQROIHOMQMEIOICROYLLLSGQTVGAPASPSSELPGTGAASSTKPLPLF 240
    |||||||
Db 181 LEEELRVLOQROIHOMQMEIOICROYLLLSGQTVGAPASPSSELPGTGAASSTKPLPLF 240
OY 241 SPIPAOTGKTATSSSSSSSSSGAEPPKQAFHLHYHPLGSHQHFVSGVGRSHKPTPAPS 300
    |||||||
Db 241 SPIPAOTGKTATSSSSSSSSSGAEPPKQAFHLHYHPLGSHQHFVSGVGRSHKPTPAPS 300
OY 301 PALPGSTQOLILASPHLAFPGTTGLLAOCIGAARGLAASPGILKPKNGSGELGYEVI 360
    |||||||
Db 301 PALPGSTQOLILASPHLAFPGTTGLLAOCIGAARGLAASPGILKPKNGSGELGYEVI 360
OY 361 SLEKPGGRHKRCFCAKVGSDSALOIHLSRSHGGERPKCNVCGNRTTRGNLKVHFRH 420
    |||||||
Db 361 SLEKPGGRHKRCFCAKVGSDSALOIHLSRSHGGERPKCNVCGNRTTRGNLKVHFRH 420
OY 421 REKYHVQMNPHRVPREHLDYVITSSGLPYGMSVPREKAEEDAGTPGGGVERKPLVASTTA 480
    |||||||
Db 421 REKYHVQMNPHRVPREHLDYVITSSGLPYGMSVPREKAEEDAGTPGGGVERKPLVASTTA 480
OY 481 LSATESLTLSTGTSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSEGSALAGVADSG 540
    |||||||
Db 481 LSATESLTLSTGTSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSEGSALAGVADSG 540
OY 541 SATRMQSLKLVSLPFWALLTNHLKSTGSPFPVLYLEPLGASPSSETSKLOOLEVEKIDRQ 600
    |||||||
Db 541 SATRMQSLKLVSLPFWALLTNHLKSTGSPFPVLYLEPLGASPSSETSKLOOLEVEKIDRQ 600
OY 601 AAVASTAGAPTTSAPAPSSASGPNQVICTRLVLSCPRALRLHYGHHGGERPKCKVC 660
    |||||||
Db 601 AAVASTAGAPTTSAPAPSSASGPNQVICTRLVLSCPRALRLHYGHHGGERPKCKVC 660
OY 661 GRAFSTRGNLRAHFVGHKTSPPAARAONSCPTCKKFTNAVYVLOOHVRYHNLGGQIPNGGSA 720
    |||||||
Db 661 GRAFSTRGNLRAHFVGHKTSPPAARAONSCPTCKKFTNAVYVLOOHVRYHNLGGQIPNGGSA 720
OY 721 LSEGGGAOENSSQOSTASPGSPPOPOSOQPSPEEEMSEEEDEEDVTDSDSLAG 780
    |||||||
Db 721 LSEGGGAOENSSQOSTASPGSPPOPOSOQPSPEEEMSEEEDEEDVTDSDSLAG 780
OY 781 RGSSEGGKKAISVRGDEEVSAGAEVATSVAAPTTVKEMDSNEKARPOHTLPPPPPPDN 840
    |||||||
Db 781 RGSSEGGKKAISVRGDEEVSAGAEVATSVAAPTTVKEMDSNEKARPOHTLPPPPPPDN 840
OY 841 LDHPQPEOGTSDVSGAMEEBAKLEGISSPMAALTQEGEGSTPLVEELNLPKAMKDPG 900
    |||||||
Db 841 LDHPQPEOGTSDVSGAMEEBAKLEGISSPMAALTQEGEGSTPLVEELNLPKAMKDPG 900
OY 901 ESSGRKACEVCGOSFPYOTALAEHQTHPKDGLFTCYFCROGFLDRATLKKHMLLAHQ 960
    |||||||
Db 901 ESSGRKACEVCGOSFPYOTALAEHQTHPKDGLFTCYFCROGFLDRATLKKHMLLAHQ 960
OY 961 VPPPAHPGPNIAITLSLVPGCSSSTIPSPGLSPFPRKDDPTMP 1002
    |||||||
Db 961 VPPPAHPGPNIAITLSLVPGCSSSTIPSPGLSPFPRKDDPTMP 1002

RESULT 3
US-09-812-633-3
; Sequence 3, Application US/09812633
; Patent No. US20020147996A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin L.
; APPLICANT: Li, Dawei
; APPLICANT: Mok, Samuel C.
; APPLICANT: Cramer, Daniel W.
; APPLICANT: Ma, Iupo
; TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
; TITLE OF INVENTION: Using Sal2
```

```
; FILE REFERENCE: 00742/066001
; CURRENT APPLICATION NUMBER: US/09/812,633
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: US 60/216,723
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 3
; LENGTH: 1002
; TYPE: PRF
; ORGANISM: Mus musculus
US-09-812-633-3

Query Match 100.0%; Score 5301; DB 10; Length 1002;
Best Local Similarity 100.0%; Pred. No. 4,6e-247;
Matches 1002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAQETGSSRLGPGCGEPERAGGDASEBHHPOVCAKCAQSFDPTEFLAHQNSCTDPPY 60
    |||||||
Db 1 MAQETGSSRLGPGCGEPERAGGDASEBHHPOVCAKCAQSFDPTEFLAHQNSCTDPPY 60
OY 61 MVIIGGDNBSNNSASAPREBGRSRQVMDTEHNPDPSSGSGPDPPTWGPERRGEES 120
    |||||||
Db 61 MVIIGGDNBSNNSASAPREBGRSRQVMDTEHNPDPSSGSGPDPPTWGPERRGEES 120
OY 121 GQFLVAAATGTAAGGGGGLILASPKLGATPLPEESTPAPPPPPPPPPPPGVGSHLNIPLI 180
    |||||||
Db 121 GQFLVAAATGTAAGGGGGLILASPKLGATPLPEESTPAPPPPPPPPPPPGVGSHLNIPLI 180
OY 181 LEEELRVLOQROIHOMQMEIOICROYLLLSGQTVGAPASPSSELPGTGAASSTKPLPLF 240
    |||||||
Db 181 LEEELRVLOQROIHOMQMEIOICROYLLLSGQTVGAPASPSSELPGTGAASSTKPLPLF 240
OY 241 SPIPAOTGKTATSSSSSSSSSGAEPPKQAFHLHYHPLGSHQHFVSGVGRSHKPTPAPS 300
    |||||||
Db 241 SPIPAOTGKTATSSSSSSSSSGAEPPKQAFHLHYHPLGSHQHFVSGVGRSHKPTPAPS 300
OY 301 PALPGSTQOLILASPHLAFPGTTGLLAOCIGAARGLAASPGILKPKNGSGELGYEVI 360
    |||||||
Db 301 PALPGSTQOLILASPHLAFPGTTGLLAOCIGAARGLAASPGILKPKNGSGELGYEVI 360
OY 361 SLEKPGGRHKRCFCAKVGSDSALOIHLSRSHGGERPKCNVCGNRTTRGNLKVHFRH 420
    |||||||
Db 361 SLEKPGGRHKRCFCAKVGSDSALOIHLSRSHGGERPKCNVCGNRTTRGNLKVHFRH 420
OY 421 REKYHVQMNPHRVPREHLDYVITSSGLPYGMSVPREKAEEDAGTPGGGVERKPLVASTTA 480
    |||||||
Db 421 REKYHVQMNPHRVPREHLDYVITSSGLPYGMSVPREKAEEDAGTPGGGVERKPLVASTTA 480
OY 481 LSATESLTLSTGTSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSEGSALAGVADSG 540
    |||||||
Db 481 LSATESLTLSTGTSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSEGSALAGVADSG 540
OY 541 SATRMQSLKLVSLPFWALLTNHLKSTGSPFPVLYLEPLGASPSSETSKLOOLEVEKIDRQ 600
    |||||||
Db 541 SATRMQSLKLVSLPFWALLTNHLKSTGSPFPVLYLEPLGASPSSETSKLOOLEVEKIDRQ 600
OY 601 AAVASTAGAPTTSAPAPSSASGPNQVICTRLVLSCPRALRLHYGHHGGERPKCKVC 660
    |||||||
Db 601 AAVASTAGAPTTSAPAPSSASGPNQVICTRLVLSCPRALRLHYGHHGGERPKCKVC 660
OY 661 GRAFSTRGNLRAHFVGHKTSPPAARAONSCPTCKKFTNAVYVLOOHVRYHNLGGQIPNGGSA 720
    |||||||
Db 661 GRAFSTRGNLRAHFVGHKTSPPAARAONSCPTCKKFTNAVYVLOOHVRYHNLGGQIPNGGSA 720
OY 721 LSEGGGAOENSSQOSTASPGSPPOPOSOQPSPEEEMSEEEDEEDVTDSDSLAG 780
    |||||||
Db 721 LSEGGGAOENSSQOSTASPGSPPOPOSOQPSPEEEMSEEEDEEDVTDSDSLAG 780
OY 781 RGSSEGGKKAISVRGDEEVSAGAEVATSVAAPTTVKEMDSNEKARPOHTLPPPPPPDN 840
    |||||||
Db 781 RGSSEGGKKAISVRGDEEVSAGAEVATSVAAPTTVKEMDSNEKARPOHTLPPPPPPDN 840
```

QY	84.1	LDHPDMEGTSVNSCAMEEBEAKLEGTSPPAAALTOEEBEGSTPLVEELNETDEAMKKDDG	900
Db	84.1	LDHPDMEGTSVNSCAMEEBEAKLEGTSPPAAALTOEEBEGSTPLVEELNETDEAMKKDDG	9000
QY	901	ESSGRKACEVCGGSPFQYRLALBEHQKTHPKDGPFLTYCYFCROGELIDRATLKHHMLAHHQ	960
Db	901	ESSGRKACEVCGGSPFQYRLALBEHQKTHPKDGPFLTYCYFCROGELIDRATLKHHMLAHHQ	9600
QY	961	VPPFAPHGPNITLSLVPGCCSSSIPSPGSLSPFRKDDPTMR	1002
Db	961	VPPFAPHGPNITLSLVPGCCSSSIPSPGSLSPFRKDDPTMR	1002

```

? RESULT 4
? US-09-988-117-1
? Sequence 1, Application US/09988117
? Patent No. US20020156039A1
? GENERAL INFORMATION:
? APPLICANT: Benjamin, Thomas L.
? APPLICANT: Li, Dawei
? APPLICANT: Mok, Samuel C.
? APPLICANT: Cramer, Daniel W.
? APPLICANT: Ma, Yupo
? TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
? TITLE OF INVENTION: Using Sal2
? FILE REFERENCE: 00742/066002
? CURRENT APPLICATION NUMBER: US/09/988,117
? CURRENT FILING DATE: 2001-11-16
? PRIOR APPLICATION NUMBER: US 09/812,633
? PRIOR FILING DATE: 2001-03-19
? PRIOR APPLICATION NUMBER: US 60/726,723
? PRIOR FILING DATE: 2000-07-07
? NUMBER OF SEQ ID NOS: 21
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 1
? LENGTH: 1005
? TYPE: PR1
? ORGANISM: Homo Sapiens
? US-09-988-117-1

```

Query Match	Best Local Similarity	88.2%;	Score 4677.5;	DB: 9;	Length 1005;	
Matches	894;	Conservative	25;	Mismatches 83;	Indels 3;	Gaps 3;
QY	1	MAOETGSSRLGPGCEPAERNGDASEEHNDPOVCAKCAQESDPTEFLAHONSCTDPV	60			
DB	1	MAHESERSRLGVPGEPAELGGSASEEDHPOVCAKCAQFDPTEFLAHONACSTDPIV	60			
QY	61	MYIIGGQENPENSASSASAPREGEHSRQVMDETHSNPPDSGSGPPDPWGERGSEES	120			
DB	61	MYIIGGQENPNNSASSAPREGEHNPPQVMDTEHSNPPDSGGSVPTDPTWGERGSEES	120			
QY	121	GQFLVAANGTAAAGGGGGLILASPKIGATPLPEESHPAPPPPPPPPPGVSCHNIPIL	180			
DB	121	GHLFLVAAGTAAAGGGGGLILASPKIGATPLPEESHPAPPPPPPPPPGVSCHNIPIL	180			
QY	181	LEELRYLDOORIHOMOMTEQICROYLLIGSQIVGAPASBSELPGTGAASSTKPLPLE	240			
DB	181	LEELRYLDOORIHOMOMTEQICROYLLIGSQIVGAPASBSELPGTGAASSTKPLPLE	240			
QY	241	SPIKPAQTKTFTA-SSSSSSSSSGAEPKQAFHLYHPLGSGHPRPSVGVGSRHKPTAP	299			
DB	241	SPIKPVQTSKTLTASSSSSSSSSGAETPKQAFHLYHPLGSGHPRPSAGVGRHKPTAP	300			
QY	300	SPALPGSTDOLILASHPLAFPGTGTGLLAOCGAAGLGAASGGLTKRNGSGELGYGV	359			
DB	301	SPALPGSTDOLILASHPLAFSTTGTGLLAOCGAAGLGAATASPLGLTKRNGSGELSTGEV	360			
QY	360	TSLELRPGGRHKRCFCAKVFSGDSALQTHLSHTGERPYKCNCVGNRFTTTRGNLKVPHHR	419			
DB	361	MGPLEKPEGRHKRCFCAKVFSGDSALQTHLSHTGERPYKCNCVGNRFTTTRGNLKVPHHR	420			
QY	420	HREKYPHNOAMPDHVPHILDVITSSGLPYGMSVPEKAEAEAGCPGGGVKKPLVASTT	479			

[illegible]

```

RESULT 5
US-09-812-471-1
; Sequence 1, Application US/09812471
; Patent No. US20020018765A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Thomas L.
; TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
; TITLE OF INVENTION: Using Mutant Viruses
; FILE REFERENCE: 00742/062002
; CURRENT APPLICATION NUMBER: US/09/812,471
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: US 60/216,723
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-812-471-1

```

Query Match	88.2%	Score 4677.5	DB 10	Length 1005
Best Local Similarity	89.0%	Pred. No. 3.3e-217		
Matches 894	Conservative 25	Mismatches 83	Indels 3	Gaps 3
QY	1	MAQETGSSRLGCGCEPAERGGDA	SEEHNPVYVCAKCAQFSDPTEFLAHONSCCTDPV	60
Db	1	MAHESESRSLGVPGEPAELGGDAS	SEEDHPVYVCAKCAQFDPTEFLAHQAACCTDPV	60
QY	61	MYILOGENPNSNSASAPARECGHS	QVMDTEHNPDPGSGSGPPDDTWP	PERGEESS 120
Db	61	MYILOGENPNSNSASAPRECGHNP	QVMDTEHNPDPGSGSVTDDTWP	GERGEESS 120

OY	121	GQFLVAAATGTAAGGGGGLILASPKLGATPLPRESTAPARPPRRPPRRPVCVGSHNIPRL	180
Db	121	GHELVAAATGAAGGGGGLILASPKLATPLPRESTAPARPPRRPPRRPVCVGSHNIPRL	180
OY	181	LLEELRVLAQQHQIQHQMOMTEOICRQVLLSLSGTVAAPASPSELPTGAASSTRKPLPLF	240
Db	181	LLEELRVLAQQHQIQHQMOMTEQICRQVLLSLSGTVAAPASPSELPTGTASSTRKPLPLF	240
OY	241	SPIKPAOTGKTTA-SSSSSSSSS GAEPKQAFFHLIYNLPSQHPRFSVGVGNASHKTPAP	299
Db	241	SPIKPVOTSKTLLASSSSSSSSSGAETPKQAFFHLIYNLPQSQHPFSAGGGRKHKTPAP	300
OY	300	SPALPGSTDOLIASPHLAFQTGTLAAOCIGAAKGLEAAASPGLLKPKNGSELCYGCV	359
Db	301	SPALPGSTDOLIASPHLAFPTSTTGILLAAOCIGAARGLEAASPGLLKPKNGSELSYGEV	360
OY	360	ISSLEKRGHNHKRCFCAYKVEGSDSAQIHLRSHTGERPYKCANCNRFTTGMKLXHFHR	419
Db	361	MGLEKPCGCKHNKRFCAYKVEGSDSALOIHLSRSHTERPYCNCVGNRFTHGMKLXHFHR	420
OY	420	HREKYPRVONMNPVRPHLDYVTTSGLPYGMGVPRPKAESEATGGCGVERKPLVAST	479
Db	421	HREKYPRVONMNPVRPHLDYVTTSGLPYGMGVPRPKAESEATGGVERKPLVAST	480
OY	480	ALSATESILTLLSTGTSJAVAPGLPTFNKFYLMAKEBKSADENTPPGSBSGAIAGVADS	539
Db	481	ALSATESILTLLSTSAGATATPGLPFAFKFYLMKAEBEKNKADENTPPGSBSGISGVAES	540
OY	540	GSATRMQLSKLYSLPEWALLTNILKSTGSPRPPLYLELPAASPFESTKILOUYEKIDRO	599
Db	541	STATULMOLSKMTSLTPWALLTNHFKSTGSPPLFCARALGASSETSKILOUYEKIDRO	600
OY	600	GAVAVASTAGAPTTSAPAPSSSA-SGPNOCCVICLRVASCPRALRLHYOGHGPERPKCK	658
Db	601	GAVAVTSAAGCAPTTSAPAPSSSASGPNOCVICLRVASCPRALRLHYOGHGGERPKCK	660
OY	659	VCGNAFSTRGMLRAHFPGHKTSAPAARONSCPTOCKFTYNAVTLQOHVNRIHLSGOIPNG	718
Db	661	VCGRAFSTRGMLRAHFGHKNSPARARONSCPTOCKFTYNAVTLQOHVNRIHLSGOIPNG	720
OY	719	SALSBCGGAADENSGESTASGPSFPPOQOSQPSPEBEKS-EEBEDEEBEEDVDSDS	777
Db	721	TALPCGCQAQENSGESTVSAGASFPQOOSQOPSPEBELISEEBEDEEBEEDVDSDS	780
OY	778	LAKGSESGEGERKAISVRDSESVYGAEEEVATSYAAPTYVKEMSNKAPQHTLPRPPP	837
Db	781	LAKGSESGEGERKAISVRDSESEAGAEVEVTYAANAAGKEMSNKETTQOSSLPPPP	840
OY	838	PNDLNHPRMCOGSTDVYSGAMEEBAKLEBGISSPMALULOBEEGSTPLVEBLINLPEAMKK	897
Db	841	PDSLDPQPMEOGSSGVYAGKEDEBKPERRSSPASALTPEBEATSYVLVELSLSQEAMRK	900
OY	898	DPEGSSGRKACEVCGSFPTOTALAEHOIKTHPKDGPLETCVFCRQGLDRATLTKKHMILA	955
Db	901	EPGESSSRKACEVCGQAPPSOAALAEHOIKTHPKDEPLTVCFCRQGFLERATLTKKHMILA	960
OY	958	HHQVPRFAPHQPONTATLSIVPGCSSSIIPSGLSFPRKDDPTJM 1002	
Db	961	HHQVPRFAPHQPONTAIALSLVPGCCPSITISGLSPFRKDDPTIP 1005	
RESULT 6			
US-09-812-633-1			
Sequence 1, Application US/09812633			
Patent No. US20020147996A1			
GENERAL INFORMATION:			
APPLICANT: Benjamin L.			
APPLICANT: Li, David			
APPLICANT: Mok, Samuel C.			
APPLICANT: Cramer, Daniel W.			
APPLICANT: Ma, Yupo			
TITLE OF INVENTION: Diagnosing and Treating Cancer Cells			
TITLE OF INVENTION: Using Sal2			

```

: FILE REFERENCE: 00742/066001
: CURRENT APPLICATION NUMBER: US/09/812,633
: CURRENT FILING DATE: 2001-03-19
: PRIOR APPLICATION NUMBER: US 60/216,723
: PRIOR FILING DATE: 2000-07-07
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 1005
: TYPE: prt
: ORGANISM: Homo Sapiens
: US-09-812-633-1

```

Query Match	Similarity	88.2%	Score 4677.5	DB 10	Length 1005
Best Local	89.0%		Pred. No. 3.3e-217		
Matches	894	Conservative	25	Mismatches	83
				Indels	3
				Gaps	3

[illegible]

```

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.94
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.97
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
OTHER INFORMATION: EST. HUMAN HIT: A0131453.1, EVALU1 1.00e-105
US-09-864-761-42953

Query Match          9.1%; Score 482.5; DB 10; Length 330;
Best Local Similarity 35.7%; Pred. No. 8,8e-17;
Matches 131; Conservative 37; Mismatches 98; Indels 101; Gaps 13;

OY 531 SATAGVADSSATRMOLSKLVTSLPMSALLTNHLKSTGSPFPYVLEPFGASPSSTKLQ 590
DB 4 SPRAADGPPASATTF-----TNPPLPLMSEQFA--KFPGGGLD--SAQASSTKLQ 52
OY 591 QLVKEDIDROGAVNAVASTASGAPFTSAPAPSSASGPNOCVILRYLSCPRALRLHYGOH 650
DB 53 QLVENIDK-----ATDPNEIILCHRYLVSQCALKMHRTH 89
OY 651 GERPEKCKVCGRAPFTNGNLRAHFVGHKTSPPAARONSCPTCKKFTNAVTLQOYHML 710
DB 90 GERPEFKCKICCGRAFTTNGNLKTHYSVHRAMPPLRVQHSCTPCKKFTNAVTLQOYHML 149
OY 711 GGGIPN-----GGSALSEGGGAOEN-----SSRQSTASGSGSPFO--POSQOPSP 755
DB 150 GGGIPIPTPVYDSYSESNESDTGSFDEKRFDDLDNSTDENMECPKESIDPTKPSADASQ 209
OY 756 E-----EMSE-----EEDEDEEEDVDYDEDSLARG-- 782
DB 210 SLSSSPLPLEMSSIALENOMKMINAGLAEOALQASLKSYENSGISRDVLTNPSSVSGDM 269
OY 783 -SRSGGEKATSVAGDSPEV---SGAEEFVATSVAAFTYKEDNSNEKAOHTLPPV--- 834
DB 270 ESGSASPALISESTSSMQLSPSNSSTQDFHKS-----PSTEEKQRAVPSEFANG 319
OY 835 -PPPPDN 840
DB 320 LSPTPVN 326

RESULT 8
US-09-864-761-43157
: Sequence 43157, Application US/09864761
: Patient No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharron G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aeomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664

```


OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EST_HUMAN HIT: AW953464.1, EVALUE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: P52736, EVALUE 0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: AW953464.1, EVALUE 0.00e+00
US-09-864-761-37152

Query Match 5.6%; Score 297.5; DB 10; Length 515;
Best Local Similarity 21.0%; Pred. No. 1e-07;
Matches 159; Conservative 73; Mismatches 235; Indels 291; Gaps 29;

QY 226 GTGAASSTYPLPLPSPDKPAQGTGTAASSSSSSSSGAEPKQAFPHLYHPLGSOHPFS 285
D 226 GTGAASSTYPLPLPSPDKPAQGTGTAASSSSSSSSGAEPKQAFPHLYHPLGSOHPFS 285
D 9 GCGA-----MPLF-----GRTKKRTLGAFS-----RPPQR-----QPVSSRN--- 40
QY 286 VGVGGRSHKPTPAPSPALPSTOLLINPHLAFPGTGTGLAQCGLGAANGLEAAASGGL 345
D 41 -GLRVELEASPAQS-GNPEETDLKRIEVLGFTVN--CGEC-----GLSPSKMTNLT 91
QY 346 KPKN-GSGE-----LGVEVYISLSEKPGGRHK-----CRCAKVPGSDALQIHLR 390
D 92 SHQRIHSGEKPYVCYCEKGFSLKSLARHOKAHSGEKPYVCREGGRGNRKSTLIHER 151
QY 391 SHTGERPYKCNVCNFTTNGNLKVHFRHREKYPHVOMNPHVPEHLDYVITSSGLPYG 450
D 152 THSGEKPYKCSGEGRGFSOKSNLIHQRTHSGEKP-----YVCRECGKFS 197
QY 451 MSVPPEKAEAEAGTPCGGVRKPLVASTTALSTATESLTLLSTGTSTAVAPGLPTFKFVL 510
D 198 QKSAVVRHRT-----HLEKTIIVCSDCGLGFSDRSNLTIS----- 232
QY 511 MKAVEPKSKADENTPPGSEGSALAGVADSGSATRMOLSKLYTSLPSWALLTNHLKSTGSP 570
D 233 -----HQRTHSGEKRYA---CKECGRCFRQRT-----LVNH-QRTHSK 267
QY 571 PFPYVLEPLGASPSSEKSLQOLVYEKIDROGAVAVASTAGAPTTSGAPADSSASGPNQCV 630
D 268 EKPYVCGVCGHFSFQNSTL-----ISHRRTHGT-----EKPYVCG 302
QY 631 ICLRLVSCPRALRLHGOHGERPFCCKVCGRAFSTRGNLRAHFVGHKTPSPARAQNSCP 690
D 303 VCGRGFSLKSHLNHRQNIHSGEKPYCKDCGRGTSQOSNL-----IRHQTHSGEKPMVCG 358
QY 691 ICQKFTNAVTLQOHVRLHGGQIP-----NGSALSSEGGAAQENSBSOSTASGPGSPFQ 746
D 359 ECGRGFSOKSNLVAHQRTH-SGERPYVCREGGRGFSHQAG----- 397
QY 747 PQSQPSPSEEMSEEEDEEEDYTDDEDSLARGSESGEKAISVAGDSEVSGAEFE 806
D 398 -----LIRHKRHSREKPYMCRCQGL-GFGNKS----- 424
QY 807 VATSVAAPTVKEMDSNEKAPQHTLPPPPPPNDLHPQMEQGTSDVSGAMEEAKLEG 866
D 425 -----ALITHKRAHSEK----- 437
QY 867 ISSPMAALTOBEGSTPLVEELNLPKAMKKDPGESSGRACEVCGGSPPTOTALAEHQ 926
D 438 -----PCV-----CRECGGFLQKSHLTLMQ 459
QY 927 THPKDPLFTCVFCROGFLDRATLKKH--MLLAHQVP 962
D 460 THGEKP-YVCKTCGRGFSLSKSHLSRHKRTTSVHHRLP 496

RESULT 10
US-09-764-864-883
; Sequence 883, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17

;; Prior application data removed - consult PAM or file wrapper
;; NUMBER OF SEQ ID NOS: 1792
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 883
;; LENGTH: 577
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-764-864-883

Query Match 5.1%; Score 270.5; DB 10; Length 577;
Best Local Similarity 21.2%; Pred. No. 2.2e-06;
Matches 144; Conservative 77; Mismatches 252; Indels 207; Gaps 26;

QY 334 RGLLEAASPELLKPKNGS-GEIYGEVYISLSEKPGGRHKRCQAKVFGSDALQIHLRSH 392
D 64 RDLDAITD--ISPKOSTHGERG-----HRCSDCKGFFLQASNFILQHRRIH 106
QY 393 TGERPYKCNVCNFTTNGNLKVHFRHREKYPH-----VQANPHVPEHLDYVITS 444
D 107 TGERPKCGCGKSYNORVHLOHVRVHTGEKPYKQYCGKAFVSSHLVQHH---SVHS 163
QY 445 SGLPYGMSVPEKAEAEAGTPGGGVERRKPLVASTTALSTATESLTLLSTGTSTAVAPGLPT 504
D 164 GERPYCGN-----ECGNFG--RHSHL-----EHLKHFREKSORCSDKRSK 204
QY 505 FNKFVLMKAVEPKSKADENTPPGSEGSALAGVADSGSATRMOLSKLYTSLPSWALLTNHL 564
D 205 NTKLSVKKRISEYSEAD-----MELS----- 225
QY 565 KSTGSPFPYVLEPLGASPSSEKSLQOLVE-----KIDROGAVAVASTAGAPTTSG 615
D 226 -----GKTQNNVSOVODPGECECFQGLDKRKGITPKKLT--GPPSKRMN 269
QY 616 -APAP-----SSASGPNQCVICLRLVSCPRALRLHYGOHGERPFCCKVCGRAFSTRGN 669
D 270 YSEVPRVHKKSTGERBPKHCKNECGKSFQSAHLIQHQHITGKEKPFCECGKSY---N 325
QY 670 LRAHNVGHKTPSPARAQNSCPICQKFTNAVTLQOHVAMHLGGQIPNGSALSSEGGAAQ 729
D 326 QVNHILQHOVRVHTGEKPYTPPLCGKAFVRSHLVQHOSSH-SGERPFCNECGKGRGRS 384
QY 730 ENSSEOSTASGPGSPFQPOSOQPSPEEMSEEEDEEEDYTDDEDSLARGSESGEGRK 789
D 385 HLAGHLRLHS-----REKSHOCRCGELFPYVSLI-----EH 417
QY 790 AISVAGDSEEVSGAEAEVATVAAPTTVKEMDSNEKAPQHTLPPPPPPNDLHPQMEQ 849
D 418 QVLHMGOKNEKNGICEEAVSWNLTVIEDKKILEOE-----QPYG- 456
QY 850 GTSVDSGAMEEAEKLEGISPMALTOEGEGTSPPLVEELNLPKAMKKDPGESSGRK--- 906
D 457 --CDICG-----KAFGYSDLI-----OHYRTHTAEKPYQODICRNVCQCSHTKQ 502
QY 907 -----ACEVCGGSPPTOTALAEHQTHPKDPLFTCVFCROGFLDRATLKKHMLLA 957
D 503 KIYSTKSHQCEGCGFLLKSHLNQHRITHTGEK--FQCEGCGANFMSCSLFAKH-LRS 560
QY 958 HHQVPPAPHGQNTATLSL 977
D 561 HERIDP-----INTLSV 572

RESULT 11
US-09-989-920-212
; Sequence 212, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sel-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P


```

PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 36336
LENGTH: 375
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC011451.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUATE = 1.00e-102
OTHER INFORMATION: SWISSPROT HIT: Q14585, EVALUATE = 1.00e-124
US-09-864-761-36336

Query Match 4.68; Score 244.5; DB 10; Length 375;
Best Local Similarity 23.78; Pred. No. 2.5e-05;
Matches 89; Conservative 40; Mismatches 141; Indels 105; Gaps 11;

QY 371 KORFCAYVGSDSALQILHRSHTGERPYKCNVCGNRFTTRGNLKVHFRHREKYPHYQMN 430
D 39 ECHGCKAPVDHSSLSKHSRSHTSKPYQCKECCGAHFLACKKMKMTPEE-----K 92
QY 431 PARYVEHLDIVITSSGLPYGKSVPEPEKAEEBAGTPGGVGERKRPVLASTALATESITLL 490
D 93 PECKEECTRAFCSSCFERRAHMKIKGTNYECKECKGK-----FSCSSSLT-- 138
QY 491 STGCTAIVAPGLTFPNKFVLMKAVEPKSKADENTPRGSSEGIAGVA-----DGSATRM 545
D 139 -----EKKRIHSGDKPYECKECKGKCAFCSSSSLSKRIHSGDKP-Y 178
QY 546 QLSKLVTLSPSMALLTNHLK-STGSPFPYVLEPLGASPSSETSKLOOLVEKIDROGAVAV 604
D 179 ECKECKGKAFSSSHLLIHIRIHNGE--KPYECKECKGKAFSESSKL----- 221
QY 605 ASTAGCAPTSAAPAPSSASGPNOCYICLRILSCGRALRLHNGOHG----- 651
D 222 --TVNGRHT-----GEKPYCKCKECKGKAYNCSSLSIHRKRIHGERPYECKECKAF 271
QY 652 -----ERPEKCKYCGRAFCSTGNLNAHVGHHTSPAPARONSCPTCOKKF 696
D 272 YLPLSTLANTHVKNOSREKPEYCKECKGKAFSCSPFRAHYRDH-----TGKIQECKECKGTF 327
QY 697 TNNAVTLQOQHVHMHNG 711
D 328 SRSSSLTEHLRTHSG 342

```

```

: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OR INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT223
: CURRENT APPLICATION NUMBER: US/09/764,864
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1792
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1061
: LENGTH: 513
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (176)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (231)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (232)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (233)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (248)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (284)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (324)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1061

Query Match          4.6%; Score 244.5; DB 10; Length 513;
Best Local Similarity 18.5%; Pred No. 3.5e-05;
Matches 131; Conservative 77; Mismatches 258; Indels 241; Gaps 21.

OY      282 HPSVSGVGSHKPTRPAPSPALPGSTDLIASP-HLAFPGTGLLAQCL--GAARGL-- 336
           || :|||              || :|||              || :|||
Db       12 HPDMGGOQSOFHSDVLGTGHEGVTPAESPLICNGCFQGPNPDILQHQIVHTGEASFQCD 71
OY      337 ---EAASPGILKKKN-----GSGELGYEVTISLSLEPRGGRIHKRF 374
           || :|||              || :|||              || :|||
Db       72 DCGKTTSQNVSYLAKNRHNSHMSEKAYQCSECGKAFFRHSDFSRHOSHSSSER---YMCNE 128
OY      375 CAAVFSGDSALOILIRSHITGERPYKCNVCNAGNETTRGNLKVNHRHREKYPHYOMNPBV 434
           || :|||              || :|||              || :|||
Db       129 CGKAFSQNSSLLKIKQKGSHMSKEPYECNCEGKAFFRRSNNLIQHRIHSXKP----- 179
OY      435 PEHLIDYITTSGLRPLYGMSVPPEKAEDEEAGTPEGGVERRKPLVASTYALSTATETLLSTGT 494
           || :|||              || :|||              || :|||
Db       180 ----YYCSECCKAFAFRS---SNLIKHNHTHTG---EKFEFCEGCEGKAFFQSANL---- 223
OY      495 STAVADGLPTFNKFVULKAAVEPKSKADENTPRGSGESALLAGVADSGSATIMQLSKLYTSL 554
           || :|||              || :|||              || :|||
Db       224 -----RKHQRYNHXXXKKPPCECNCGKPFRRXS----- 249
OY      555 PSWALLTNHLKSTGSPFPFVVLEPLGASPSETSKLOQ--LVEEKIDROGAVALASTA-SGA 611
           || :|||              || :|||              || :|||
Db       250 ----NLTKHNHVHTGE--KPYKCSDOCKAFSGOSSLLIQNHXIHNGEKPHVNCVKGFATSYS 304
OY      612 PITTSAAPRSSASGAPNOCVLCRLVLSCPRALRLAHYGOGGERPFGKVCVGRAFSTRGNLR 671
           || :|||              || :|||              || :|||
Db       305 SVLRKHQIILHTGERPYKSCXCKGKAFSHSALLIQHGTVHTGDGRPYACHCEGKTFGRSSNL- 363
OY      672 AEFVGIKTESPAARAONSCPILOCKFTFNNAVTLQOOHVBMHJGGDIQNGSALSSEGGAQDEN 731
           || :|||              || :|||              || :|||
Db       364 ----ILHQRVHTGEKPYRCTSECGKTFSGOSTLIOHORINNG----- 400

```

QY 732 SSEOSTAGPGSFPOQSOQSPSEEMSEEEDEEDVTDSDSLAGSGSESGEKAI 791
Db 401 ----- 400
QY 792 SVKGDSEEVSGAEVEEVATSVAAPTTVKEMDSNE--KAPQHTLPPPPRPNDLHPQMEQ 849
Db 401 -----LKPHECNCQCGKAFNRS-----SNLIHHQKVT 427
QY 850 GTSVDVGAMEEAKLEGISPMALQIPEGEGSTPLVEELNLEPMKKDGESSGRKACE 909
Db 428 GERYPT-----CVECGKGF-----QSHLIOHOIHTGERPK--CS 463
QY 910 VCGSFPOTALLEHQHTPKDGPLFTCVCFRCROGFLDRATLKKHML 956
Db 464 ECGAFSGRSVLIQHRIHTGVKP-YDCAACGKAFSGRSKLINHQLI 509
RESULT 15
US-09-864-761-33476
; Sequence 33476, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33476
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens

; FEATURE:
; OTHER INFORMATION: MAP TO AF118808.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3
; OTHER INFORMATION: SWISSPROT HIT: P52742, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE885790.1, EVALUE 1.00e-116
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUE 1.00e-115
US-09-864-761-33476
Query Match 4.6%; Score 242.5; DB 10; Length 439;
Best Local Similarity 22.0%; Pred. No. 3.7e-05;
Matches 117; Conservative 42; Mismatches 201; Indels 173; Gaps 19;
QY 227 TGAASSTKPLPLPSPIKPKNOT-----GKTTASSSSSSSSGAPRKOAFNHLVPLGSDHP 283
Db 3 TKGESQNTDLSR--KPLISEQTVILGKTRPLGRIDENNE---TKQSF-----CLSD 48
QY 284 FSYGVGVSRSHKPTAPAPALPGSTQDLIASPHLAPP-GTTGLAAQCLGAARGLEMAASP 342
Db 49 NSV-----DIREVYVLSQMP-----LTPHOAVPSGERPYMCVGC-GKCFG-----88
QY 343 GLLPKNGSGELGYGEVYSSLEKPDGRHKRCFAKVPFGSDALSQIHLRSHHTGERPKCV 402
Db 89 -----RSSHLIOHORIHTEGKP--YVCSYCGKAFSSSVLSKRRRIHTGKPYECNE 138
QY 403 CGNRTTGNLKVHFNHREKXYHQMNPHRYBELIDVYITSSGLPYGMSVPREKAEDEA 462
Db 139 CGKAFRVSSDLQHNKIHTEGKPYECLECRKAFTQLSHLIOHORIHTEG-----186
QY 463 GTPGGVEKRPVASTTALSTESITLTS-----TGTSTAVALGPTFNKFLMKAVEPK 517
Db 187 -----ERYVCPRLCGKANHSTVLRSQHVHVG-----EKP 217
QY 518 SKADENTPPSGESGALAGVADSGATRMQLSVLSPSW-----ALLTNHL 564
Db 218 HRCNE-----CGKTFESVKKRTLQHORIHTEGNEPYTSGECCGKARSRLVLIQHN 266
QY 565 KTGSGFPE-----PYVLEPLASPESTSKLQQLVEKIDR 598
Db 267 VHTGEKPYECSECGKTESHRSTLMNHERIHTEKRPACYECGAFVQHSHL-----IQH 320
QY 599 QGAVALVASTAGAPTTAPAESSASGPNQCVICLVLSLCPRALRYLHGQNGGERPEKCK 658
Db 321 QRV-----HTGEKRYVCGEGCGHAFARSRLIGHRIHTEKPEKPOCT 361
QY 659 VCGAEFSTRGNLRAHFVGHKTSPPARAONSCPIQCKFTFNATVTLQOHVRRHNG 711
Db 362 ECGKAFSLKATLIVLIRTH---TGKPYECNCGKAFSQQSVLIQHORIHTEG 410
RESULT 16
US-09-925-301-1357
; Sequence 1357, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1357
LENGTH: 580
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (3)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (526)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1357

Query Match 4.6% Score 242; DB 10; Length 580;
Best Local Similarity 23.7%; Pred. No.5.2e-05;
Matches 105; Conservative 54; Mismatches 178; Indels 106; Gaps 16;

QY 322 TGIILAAOCIGAARG-LEMAASPGLKPKNGSG-----ELGYGEVSS-----LEK 365
DB 182 TGEKAFDCSDSGKSFIMHSHLOGHLRTNGESLHEMKECGRGIHSTDLAVRIOTHRSK 241
QY 366 PGRHRCRFCAKYPGSDSALOILHRSHTGERPKYKCNVCNRRFTTGNLKVHFRHREKYP 425
DB 242 P---YKCEKCGKGFERYSAYLNIHMGTHGTGDNPECKEKGKAFRSCOLQH-----RKT 292
QY 426 HYQMNHPVPEHLDYVITSSGLPYGMSVPRPEKAEAEAGTPGGVEKKPLVASTTALSATE 485
DB 293 HNGEKRYKCKDCGRATFVSSCLSOHMKIHVGKPYECKEG-----IAFRSSOLTE 344
QY 486 STLLSTGTSTAVAPGLPTFNKFLMKAVEPK--SKADENTPEGSESAI-AGV-----A 537
DB 345 HKK-----FHTADP-----FECKICKSFRNSSCLSDHFRHTGIKPKCK 386
QY 538 DGSATRMQLSKVITSPSMALLTNHLKSTGSFPPPYLEPLGASSESKL-----589
DB 387 DCGKATTONSD-----LTKHAR-THSGERPVECKEKGKAFARSRLSEHRTHT 434
QY 590 -QOLVEKIDROGAVAVASTASG-----APTTSAPARSSASGP 626
DB 435 GKPFECVACGKAFATSSNLGHLRIHNGEKPEFCELECKATTHSSSLNNHMTSHAKK 494
QY 627 NCCVLCRLVSCPRALRLHYGOHGGERPKKVCGRAFSTRGNLRAHFVGHKTSPPARAQ 686
DB 495 FTCMEGKAKFKPTCVNLHMRIHNGEKPKYKXCKGKSFYSNSFQLHERTH-----TGEKP 550
QY 687 NSCPICQKKFTNAVTLQOHVBMH 709
DB 551 YECKECGKAFSSSSSFRNHERRH 573

RESULT 17
US-09-735-367B-3
Sequence 3, Application US/09735367B
Patent No. US20020151477A1
GENERAL INFORMATION:
APPLICANT: Gustafsson, Jan-Ake
APPLICANT: Caixa, Francoise
APPLICANT: Antousson, Per
TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR
FILE REFERENCE: 102093-100
CURRENT APPLICATION NUMBER: US/09/735,367B
CURRENT FILING DATE: 2000-12-12
PRIOR APPLICATION NUMBER: US 60/174,544
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PastSeq For Windows Version 4.0
SEQ ID NO 3
LENGTH: 2005
TYPE: PRT
ORGANISM: mammal
US-09-735-367B-3

Query Match 4.6% Score 242; DB 10; Length 2005;
Best Local Similarity 19.5%; Pred. No.0.00019;
Matches 225; Conservative 105; Mismatches 380; Indels 444; Gaps 46;

QY 19 AERGDASEEHHPYCAKCAQAFSDPTEFLAHQNSCTDP-----PYMVI-- 64
DB 777 AMGNASAGNHFSGHGMFNAFPGCAPN--GNQMSGQNGFVNNKDVTLSPLLVNLQ 834
QY 65 -----GGOENPSNNSASAPRPGHSRQYMDTEHSPNPDSSGS-----PPDPTWG 111
DB 835 SDISAGHFVYNNKQNNNTNANKPKKKKNSQODLTPDTRPAGLEBADQPLP--G 892
QY 112 PERGESSGQFLVAATGATAGGGGLIASPKITATPL--PPE--STRAPPPPP-- 163
DB 893 EOGTSLDNGCPKLEDFSRPPC-----YPSQPVGEORPLQOMPPQLMQHVAAPPQPPQQ 946
QY 164 -----PPPPVGVSGHLNIPLELRLVLOQROIHQMQMTEQICROVLLIGSGQT 214
DB 947 PQDPLPQQQQPPPP-----SOPQSOQQQQQQQQMMMLMQODPKS 987
QY 215 VGAPASPELPGTGAASATKPLPLFSPIKPAQTGKTAS--SSSSSSSGAEPPKQAF-F 272
DB 988 VRLPVSONVHPPRG-----PLNPDQRMPMQOOSGSVPYVWVSLQCPASVPPSPDKQMPM 1041
QY 273 HLYHPIGSQHPFSVGVGRSHKPTPAPAPALPGSTQDLASPHLAPGTGTGLAAQCLGA 332
DB 1042 PVNTPLGS-----NSRKNVYQESFQNPSS-----SP-----LLEMASLPE 1076
QY 333 ARGLEAASBP-----LLKPKNGSGELGYGE-----VISLEKPGG 368
DB 1077 ASGSEAPSVPGGRNMPSHVVLPOQLMWTGPKPGPSPLSATQGATPQOPVWVSLPSHG 1136
QY 369 RHKCRFCAKYPGSDSALOILHRSHTGERPKYKCNVCNRRFTTGNLKVHFRHREKYPHQ 428
DB 1137 HH---FPNVAAAPQTQSRPKTPNRASPRPY-----YQTP 1167
QY 429 MNHPVPEHLDYVITSSGLPYGMSVPRPEKAEAEAGTPGGVEKKPLVASTTALSATE 488
DB 1168 NNRPPSTE-----PSEISLSPRR-----LNASIAGLPPPOINT 1200
QY 489 LSTGTSTAVAPGLPTFNKFLMKAVEPKSKADENTPGESESAIAGVADSGSATRMQLS 548
DB 1201 PL-----PPRPNLNRGFDQGLNPTT-----1221
QY 549 KLVTSLSMALLTNHLKSTGSFPPPYLEPLGASSETSLOQVLEKIDROGAVAASTA 608
DB 1222 -----LKAIGAPSNLTNMPNFATPQTHKLDVYVNSGKQ-----SN 1259
QY 609 SGAPTSAPAPSSASGPNOCVLCRLVLCPRALRLHYGOHGGERPKKVCGRAFSTRG 668
DB 1260 SGATKRASPS-NSRRSSP-----GSSR-----1280
QY 669 NLRAHFVGHKTSPPAARQNS-CP-ICOKKFTNAVTLQOHVRN-----HLGG 712
DB 1281 -----KTPSPGRQNSKAPKLTLLASQTNAAAL-ONVELPRNLVSPPTPLANPPVG 1330
QY 713 QIPNGS-----ALSEGGAGQENSSEGSTASGSPSPFPP--QSOQPSSEEMSEE 761
DB 1331 SFPNNSGLPNQNSTVSAVAAVGVEDNKESLNV-----PQSDSQNSQSRKEQVNIEL 1383
QY 762 EEEDEEEEDVTDSDSLAGNSESSEGEKAI SVRGDSEEVSGAEFEVATVVAAPTYKEMD 821
DB 1384 KAVPAQEVKMYVEEDSKKQGPSPDKLPVEEENKNLVSPAMRE-----APTSLSQL 1437
QY 822 SNEKAPQHTLPPP-----PP-----PPDN-----840
DB 1438 DMSGAPNVTIKPPGLDLEVTYPVAVSGEDLKAKAVIPTLQDLSSEKSPSNTLPHSNEL 1497
QY 841 ---LDHPQMEGQTS---DYSGAMEEBAKLEGISSPMALQOEGGCTSPPLYBELNLPRA 894
DB 1498 CSSLVHPELSEVSSNAVAPSPYVMSRPVSSSISTPPLPNOITVFTVTSNPITTSANTSA 1557
QY 895 MKKDPGSSGKACAEVCGGSFPT--QTALDEHQKTHPKDGLFTGCVFCGQ-----943

```
Db 1558 L-----PTHQSALMSTVVTMPNAG---SKYVNSGQSAQSNAR 1554
Qy 944 -----FLDRATLKHHMLAHQVPPFAP-----HGPONI--ATLS 976
Db 1595 PQTTPYFINSSIIQVWKGSPSTIPAAPLTNTSGLMPPSVAVGRLHPQNIKFFSSAP 1654
Qy 977 LVPGCSSSIPSPGL 990
Db 1655 VPPNALSSSPAPNI 1668

RESULT 18
US-09-735-367B-2
; Sequence 2, Application US/09735367B
; Patent No. US2002015147A1
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Jan-Ake
; APPLICANT: Caltra, Françoise
; APPLICANT: Antonsson, Per
; TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR
; FILE REFERENCE: 102093-100
; CURRENT APPLICATION NUMBER: US/09/735,367B
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/174,544
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2063
; TYPE: PRT
; ORGANISM: Human
US-09-735-367B-2

Query Match 4.6%; Score 242; DB 10; Length 2063;
Best Local Similarity 19.5%; Pred. No. 0.00019;
Matches 225; Conservative 105; Mismatches 380; Indels 444; Gaps 46;

Qy 19 AERGDASEHHPOVCAKCAQPSDPTFLAHQNSCCTDP-----PVMVII-- 64
Db 835 AMQNSASQNHFSGHGMSFNAFSGAPN--GNQNSGQNFGEFVYKNDVTLTSPLLVNLQ 892
Qy 65 -----GGOENPNSSSASSAPREGHRSQVMDTEHSNPPDSSSG-----PPDPTWG 111
Db 893 SPISAGHEGVNNKQNTNANKPRKKKKKNOQDLNTPDTTRAGLEADQRPPL--G 950
Qy 112 PERRGESSGQFLVAATGTAAGGGGLILASPKLGATPL--PPE--STRAPPPPP-- 163
Db 951 EGGISLDSNGSPKLPEFSNRRPG-----YPSQVYEQRPLOQMOLMNVAPRPQPOQ 1004
Qy 164 -----PPRRPGVSGSLNPLRLIEELRVLDQROIHQOMTEQICROVILLGSLGOT 214
Db 1005 POPQLPQQQPPPP-----SOPQSQQQQQQQQMMMLMMQODPFS 1045
Qy 215 VGAPASPELPGGAASSTKPLPLFSPIKPAQTGKTAS--SSSSSSSSGAPPRQAF-F 272
Db 1046 VLLPVSONNHRPG-----PLNDSQRMHQSGSVVVMVSLQGPASVPRPBDQRMHM 1099
Qy 273 HLXHLPGSHRFSVGVGRSHKKTTPADSPALPGSTDOILASPHLAFPGTTGLLAQCLGA 332
Db 1100 PVNTPPLGS-----NSRKMYQGESPPQSS--SP-----LAEMASLPE 1134
Qy 333 AGLLEAAASFG-----LLKPKNGSGELGYGE-----VSSLEKPPG 368
Db 1135 ASGSEAPSVYGGGNMPSHVLLPONQLMTGPKGSPPLSATQCATPQOPPVNSLPSSSHG 1194
Qy 369 RHKCRCAKAFGSDSALQIHLRSHTGERPYKCNVCGNRFETTGKLVKPHRHREKYPHYQ 428
Db 1195 HH-----FPNVAAPQTGRPKTPRNASPRPY-----YPPPT 1225
Qy 429 MNDPVPENHLDYVTSSGLPYGMSVPRPKAEBEACTPGGCVERRKPLVASI TALSATSESL 488
Db 1226 NNRPSPTE-----PSEISLSPER-----LMSIAGLFPPOINI 1258
```

```
Qy 489 LLSTGTAVAAPGLPTEFNKPEVLKAVEPKSKADENTPPGSEGSALAGVADSGSATRMOLS 548
Db 1259 PL-----PPRNLNRRFPDQGLMPTT----- 1279
Qy 549 KLVTSIPSWALLTNHLKSTGSEFPFVYLEPLGASPSSETSKLQOLVEKIDRGAVAVASTA 608
Db 1280 -----LKAIQAPSNLTMMNPSNFATPQTHKLDYSVVVNSGKQ-----SN 1317
Qy 609 SGAPTTSAAPSSASGPNOCVICRLVLSCPRALRLHNGGGEFFPKCYGGRAPFTSG 668
Db 1318 SGATKRAAPS-NSRRSSP-----GSSR----- 1338
Qy 669 NLRAHFVGHKTSAPAAKONS-CP-ICQKFTNAVTLQOHVR-----HLGG 712
Db 1339 -----KTPSPGQNSKAPKLTLASOTMALL-QNVELPRVILSPTPLANPPVG 1388
Qy 713 QIPNGGS-----ALSEGGAAQENSSEOSTASGPGSFPP-----OSQOPSPDEEMSEE 761
Db 1389 SFPNNSGLNPQNSTVSVAAGVGEDNKESLNV-----PQSDCCONSQKREQVNIEL 1441
Qy 762 EEDDEEEEDVTDSDSLAGRSESGEKAISVRGDSSEVSGAEEEVATVVAAPTIVKEMD 821
Db 1442 KAVPADQVKMNVPRDQSKDQSPDPNKLPSVEENKNLVSPAMRE-----APTSLOQL 1495
Qy 822 SNEKAPQHTLPP-----PP-----PPDN----- 840
Db 1496 DNGAPRVNTIKPGLTLDLEVTPPVYSGEDLKKASVITPLQDLSSEKSPNSLNLPHSNEL 1555
Qy 841 --LDHPQMEQGT--DVSGAMEEELKLGISSPMALTOEGESTSPVLELWLEPA 894
Db 1556 CSSLVHPHELSEVSSNVAFSIPVMSRPVSSSISTPLPNOITVFVTSNPITTSANTSA 1615
Qy 895 MKKDPRGSSGKACQVCGSPPT--QTALDEHQKTHPKDGLFTVCFCRQO----- 943
Db 1616 L-----PTHQSALMSTVVTMPNAG---SKYVNSGQSAQSNAR 1652
Qy 944 -----FLDRATLKHHMLAHQVPPFAP-----HGPONI--ATLS 976
Db 1653 PQTTPYFINSSIIQVWKGSPSTIPAAPLTNTSGLMPPSVAVGRLHPQNIKFFSSAP 1712
Qy 977 LVPGCSSSIPSPGL 990
Db 1713 VPPNALSSSPAPNI 1726
```

```
RESULT 19
US-09-764-864-948
; Sequence 948, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 948
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (335)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-764-864-948
```

```
Query Match 4.6%; Score 241.5; DB 10; Length 338;
Best Local Similarity 24.2%; Pred. No. 3.1e-05;
Matches 88; Conservative 38; Mismatches 154; Indels 83; Gaps 11;
```

354 LGYGEVYISLEKPGCRHRCRCARVGVSDSALOIHLSHNGRPPYKCNVCNRETTGNTL 413
Db LGNGI-----KPF-----HSCPRCSKLPKOPSHIQTHLLTHOGTPPHKCOVCHKATQTSHL 85
OY 414 KYHFHNRREKYPHVQMPDHPVPEHLDYITSSGLPYGMSVPRKAEBAEATPGGVEKRP 473
Db 86 KRMLHLHSEKPY-----SCH-----FCGRGFAPSEIKAEHVEKHSRCHVCVE--- 130
OY 474 LVASTALASATESLTLSTGTSTAVAPGLPTFNKPYLMKAVEKSKADEMTPPGSEBSAI 533
Db 131 -----CGIDPSTLTQIKRHLASHOGPTLYCLE----- 158
OY 534 AGVADSGSATRMQLSKLVTSILPSWALTNHLKSTGSFPPYVLEPGLASPSETSKLOOLV 593
Db 159 ---CDKSEHYRSQIQ-----NHMLKHQNVK--PVCOTEGCMERSQIHLKQ-- 199
OY 594 EKIDRCGAVAVASIASAPPTSAAPS-----SSASGRNQCVCICLVLSCPRALRLHYGO 648
Db 200 HSLHTHGVKEKCEVCGRREFILOANMKRMLHTHTSVRYQCHICEFTFYQKQTLKTHIV 259
OY 649 HCGERPCKVCGRGFSTRGNLRAHFVGHKTSPPARAONSCPTCKKFTNAVTLQCHVVM 708
Db 260 HBPVKCFKCKVCKCSFMRMTNL-----LGHMHLHASKPCKPCYCSKFNKMLKMLSRHMKV 315
OY 709 HLG 711
Db 316 KHG 318

RESULT 20
US-09-864-761-35612
Sequence 35612, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wenliang
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecolica-x-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-10-03
PRIORITY APPLICATION NUMBER: GB 24263,6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 35612
LENGTH: 1325
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000511.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98
OTHER INFORMATION: EXPRESSED IN HB1100, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EST. HUMAN HIT: AM867076.1, EVALUATE 1.00e-103
OTHER INFORMATION: EST. HUMAN HIT: AM867076.1, EVALUATE 1.00e-103
OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUATE 5.00e-63
US-09-864-761-35612

Query Match 4.5%; Score 239.5; DB 10; Length 1325;
Best Local Similarity 20.6%; Pred. No. 0.00016;
Matches 204; Conservative 106; Mismatches 429; Indels 249; Gaps 40;

OY 3 OEQSSSRILGCGPGEPRERG-DASEHHNPQVACCAQFSDPTFLAUNSCCT--DPP 59
Db 242 QETISANELTQSLAEPRHGGRTANENTPSP-----AEPTENRETTANENTTLSAEPT 236
OY 60 VVVIIGQENPSSSSASAPRPEGRSHSQVM-----DTESS--NPPDSGSGP 105
Db 297 -----ENERTANENTTAPPAPTENRETTANENTTLFPAEPTEHGERTANENTTSP 349
OY 106 PPPWPERGERGSSGQFLVAATGAAGCGGLIASPKLGA-----TPRPE- 153
Db 350 AEP-----EHGERTANENTTSPAEPTHEGERTPPANCKTSSSAESTHEGERTPLANEN 405
OY 154 STPAPPP-----PPPPPGVSGHNIIPILBELAVLQOQIOMQMTQEI 201
Db 406 TTPSPAEPTENRETTANENTTSPAGP-----TENRETTANE----- 442
OY 202 CROVLLISLGQTVGAPASPSSELPGTGAASSTKPLPLPSPIKAQOTKRTASSSSSSSS 261
Db 443 -----KTLSPVEPTENRETTANENTTSP-----SPAEPTENGQRTFPANEKITS 486
OY 262 SGAEPKQAFPHLHYHPLGSHPSVGVGRSHKPTPA-----PSPALPS 306
Db 487 SGAEPTEH-----GETPLA-----NENTTSPAPTEHRETTANENTTSPAP-- 531
OY 307 TDOLASPHLAEPTGTGLAAOCLGAARGLAASPGILKP-KNGSGELGYGEVYISLEK 365
Db 532 TENGDRTPLANEKTL-----PLAEPTENQORFPANEKTTSSSA 571
OY 366 PGRKIKCR--FCAKYGVSDSALOIHLSHNGER--PYKNCVGNRETTGNLKYHF-- 417
Db 572 EPTHEERTPLANENTTSPAEPTENRETTANENTTSPAPTEHRETTANENTTLFPAE 631
OY 418 ---HHREKYPHVQMPDHPVPEHLDYITSSGLPYGMSV--PEKAEBAEAGPGGVEK 472
Db 632 PTEHRETTANENTTSPAEPTEN-----GORTPRANEKTTSSPAEPTHEGERT 679
OY 473 PLVASTALASATESL--TLSTGTSTAVAPGLPTFNKPYLMKAVEKSKADEMTPPGSEG 530
Db 680 PLANENTTSPAEPTENRETTANENTTTPPAEPTEN-----ERTANENTTTPSP-- 728
OY 531 SAIGVADSGSATRMQLSKLVTSL-----PSWALTNHLKSTGSPEP 572

Db 729 ---AQPTEGDRTPLANEKTTPSLAEPTEGCKRTPPANECTTSSSAEPTEHAERT----- 780
QY 573 PYVLEPIGASSETSRLQOLV--EKIDROGAVAVASTAGAPTTAPAPSSSASGPNQCV 630
Db 781 PLANENTTSSPAEPTEGERTPLANETTPQPAEPTEGERTANECTTPPAEPTEGERTANECTTP 840
QY 631 ICLAVLSPRALRLHYGQHGGERPF--KKVCGRASFSTRGNLAHFVGHKTSPPARAQNS 688
Db 841 NENTTSLPAEPT-----EHEEMTPLANEKTTPSLAEPTEGERTTPNEKTTP----- 888
QY 689 CPICOKKFTNNAVTLQOHVHMLGQI--PNGSALSCEGGAQENSS--EOSTASGPGSFP 745
Db 889 -----SSAEPTEGERTPLANETTPSRAPTEGERTANECTTPPAEPTEGERT 940
QY 746 QPQSOQSPPEEEMSEEEDEEEDVDEDSLAGSGESGGEKAIVRGD----SEEVVS 801
Db 941 VNEDTPSSAEPTEGERTPLANENTTSTPTESTEGERTANECTTPSPAEPTEGERTP 1000
QY 802 GAEERVATSVAPTVKEMDSNEKAFQHTLPPRPDPNDLDHPQ--PMEOGTSDVSGAMEE 860
Db 1001 SANEXTTSPAPTEHEEM--TPSANENTTSPVPKPTHEGKTTLANEKTTPSPEGTEH 1058
QY 861 EAKL-----EGISSPMALTOEGEGTSP 884
Db 1059 GAKTTSANEKTIPLAKPTEHGERTTSP 1086

RESULT 21
US-09-864-761-36708
; Sequence 36708, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00660
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36708
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006213.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EST_HUMAN HIT: A0123664.1, EVALU2 2.00e-74
; OTHER INFORMATION: SWISSPROT HIT: Q02386, EVALU2 0.00e+00
US-09-864-761-36708

Query Match 4.58; Score 238; DB 10; Length 342;
Best Local Similarity 26.38; Pred. No. 4.7e-05;
Matches 94; Conservative 28; Mismatches 121; Indels 114; Gaps 14;

QY 364 EKPGGRHKRCRCACVFGSDSALQIHLRSHTGGRPYKCNVCGNRFTRGNLKVH--FRRHR 421
Db 82 EKP--YKGEHCYKSFSSSHLQINORAHAGKPKYCKEKCNDNAFRFSSLQAHQVHSRA 138
QY 422 EKYPHVOMNHPVPEHLDVYTS--SGLPYGNSVP---PEKAEEDAGPGGGVERKPLV 475
Db 139 KSYTN-----DASYRFSQSRSHLPHQRVPTGENPYKYEEDGRRNVKS----- 181
QY 476 ASTTALSATESLTLISTSTAVAPGLPTFNKFVLMKNAEPKSKADENPPGSEGAIA 535
Db 182 -----SHQAPRLIVHTG-----EKPYKCE-----CG 203
QY 536 VADGSAATRMQLSKLVTLSPWMLTNHLK-STGSFPFPYVLEPLGASSETSRLQOLVE 594
Db 204 V--GFSQR-----SYLOVHLKVHTGK--KRYKCEGCKGSFWRBRLO----- 241
QY 595 KIDROGAVAVASTAGAPTTAPAPSSSASGPNQCVICTRLVSLCPALRLHYGQHGGERP 654
Db 242 -----AHERIHTEGKPYKCNACKSFSYSSHLNHCHIRHTEGK 280
QY 655 FKCYVCGRAFSTRGNLAHFVGHKTSPPARAQNSCTICQKFTNNAVTLQOHVHMLG 711
Db 281 YKCECGKGFVSGHLOAHQISH---TGEKPYKCECGKGFCRASVLDLDHOGHTG 333

RESULT 22
US-09-864-761-37065
; Sequence 37065, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359

```
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 37065
/ LENGTH: 551
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC005261.1
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
/ OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
/ OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
/ OTHER INFORMATION: SWISSPROT HIT: P51522, EVALUOE 0.00e+00
/ OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUOE 1.00e-105
/ OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUOE 1.00e-105
/ US-09-864-761-37065

Query Match 4.5%; Score 237; DB 10; Length 551;
Best Local Similarity 25.3%; Pred. No. 8.5e-05;
Matches 95; Conservative 39; Mismatches 157; Indels 84; Gaps 14;

QY 359 VISSLKPGGRHRCRCACVFGSDSALQIHLKSHTEGRRPYGCNVCNFRFTTGKLNKLVHFN 418
D 175 MIHTGERP---TECEMGCAFNRKRSYLTQHQRHISGEKPYRCNECGKAFTHRSNFWLHNR 231
QY 419 RHREKYPHYQMNPHRYPEH-----LDVITSSGLPYGMSVPEKAEKEAG----- 463
D 232 RHTGKSPVCTCGGQVFRHRRPGFLRHVYVHNGENPYECL-----ECGKVFKHRSYLM 283
QY 464 -----TPGGGVERKPLVASTATASLESLLSTGTSTAVAGBLPTFNKFLMKAVEPKS- 518
D 284 WHOOTHNG---EKPYECSECGKVFLESALIH---HYVIHNGEKPFLECGKATNHRXY 337
QY 519 -KADEMPGSGSALAGVADGSGATRMQLSLVTSPLSNALLNNHLS-IGSPFPPIYL 576
D 338 LKRRIORHTGKRPVCS-----ECGKAFTHCSTFL---HKRAHNGEKPFEE--C 381
QY 577 EPLGASPESTKLOLVEKIDRGAAVAASATGAPTTASAPASSASGPNOCVCLRLV 636
D 382 KECKKAVS-----NRKDLIRHPSIHNG-----EKPYECVCGKAF 416
```

```
QY 637 SCPRALRLHYGQGERPEKCKVCGRAFSTRGNLRAHFVGHKTSPPARAONSPICQKKE 696
D 417 TRMSGLTRKRRKHSGEKPYECGCKSFCWSTNLJHAIH-----TGEKPYKCECGKAF 472
QY 697 TNVATLQOHVPMHIG 711
D 473 SRSSSLTQHQRHNG 487

RESULT 23
US-09-816-669A-14
/ Sequence 14, Application US/0981669A
/ Patent No. US20020137019A1
/ GENERAL INFORMATION:
/ APPLICANT: GARABEDIAN, Michael
/ APPLICANT: TANEJA, Samir
/ APPLICANT: HITTELMAN, Adam
/ APPLICANT: MARKUS, Steven
/ TITLE OF INVENTION: METHOD FOR SCREENING TRANSCRIPTIONAL COREGULATORY PROTEINS OF
/ TITLE OF INVENTION: TRANSCRIPTION FACTORS, AND ANDROGEN RECEPTOR TRANSCRIPTIONAL
/ FILE REFERENCE: GARABEDIAN-1.1A
/ CURRENT APPLICATION NUMBER: US/09/816,669A
/ PRIOR APPLICATION NUMBER: 2001-03-26
/ PRIOR FILING DATE: 2000-08-15
/ PRIOR APPLICATION NUMBER: 60/225,618
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: 60/191,768
/ PRIOR FILING DATE: 2000-03-24
/ NUMBER OF SEQ ID NOS: 20
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 14
/ LENGTH: 2783
/ TYPE: PRT
/ ORGANISM: Human
/ US-09-816-669A-14

Query Match 4.5%; Score 236.5; DB 10; Length 2783;
Best Local Similarity 20.0%; Pred. No. 0.00048;
Matches 239; Conservative 116; Mismatches 399; Indels 443; Gaps 57;

QY 34 CAKCAQFSDPTFEIHAQNSCC---TDPVAVIIGQENPSSSSASAPRBSGSQVM 90
D 1416 CKKSLVFGIIFDLIKHQKLYCDEDEE-----GQDSQNSDS-----M 1455
QY 91 DEHSNPPDSGSGPDPPTW-----GPRKEESSGQFLVAA-----TGTAGGGG 136
D 1456 DAMEIITPTSSSCSTPSPQAYSAAPASANTASSAFLOLTAEEELATFNSKTEAGD-- 1513
QY 137 GLIASPKLGATP-LPEESTPAPRPPPPPPGVSGHNLNPLLELRVLAQOQRIHQM 195
D 1514 -----EKPKLAENASAPNOQEOKOGPKP-----ELQ--QOEOPEOK 1549
QY 196 QMT-EOICROVILLGSLGO-TVGAPASPSLEPGTGAASSTRPLRPLFSPDKPAOTGKTTA 253
D 1550 TMTPOQKLPQLVSLPSPLPQPPQAPPPQCLP-----QSSPSPQSLHPLKXLTHT----- 1600
QY 254 SSSSSSSSSGAEPP-----KQAF-----FHLNPLGQHPSPVSGVGSRSKP 295
D 1601 ---STPQOLANLPPQILPYQCDCKLAFSPFEHWOEHQOLHFLSAONF----- 1646
QY 296 TPAPSPALPGSTQOLIASPHLAFPGTGLLAOCIGAA----- 333
D 1647 -----IHQFPLDRSLDMPFMFLDPSPNPLASQLLSGAIPOIPASSATSPTPTSTMTNL 1700
QY 334 -RGLE--AAASFGLKPKNGSG----- 352
D 1701 KRKLEEKASASPG-----ENDSGTGGEPPQDRKRLRTTITPQLEITLYQKYLDSNPTRMK 1756
QY 353 -----ELGYGEVISTL-----EKPG-----GRKRCFCAVFGSDSLQIHL 389
D 1757 LDHIAHEVLKRRVVOVWFQNTRARERKRGQFRAVGAQAHRRCPFCRALFRAKTALTEAH 1816
```



```

PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 35894
LENGTH: 340
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AF166490.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 9.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3
OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUE 1.00e-91
OTHER INFORMATION: SWISSPROT HIT: P52742, EVALUE 1.00e-118
US-09-864-761-35894

Query Match 4.4%; Score 234; DB 10; Length 340;
Best Local Similarity 23.1%; Pred. No. 7.2e-05;
Matches 86; Conservative 44; Mismatches 133; Indels 110; Gaps 12;

358 EVISLEKPGGRHKCFCAKVFSGDSALQIHLRSHTGERPKVCNVCNRTTGNLKVHF 417
50 EMIHNGEKP--HGCILCGKAFTHCSDLRKHERHTL-----GDRPYCGLCGKAFSSNLRHE 106
418 HRH-REK-----YHVCNPNHPEVREHLDVYITSSGLPYGKSVPPER-----AEE 460
107 MIHTRKKQOICHLGKAFTHCSDLRKHERHTL-----GDRPYCGLCGKAFSSNLRHE 160
461 EAGTGGGVERKPLVASTALSTESTLTLSTGTSTAVAPLPTNKVPLMKRAVEKS-- 518
161 HERTING---EKP-----*ECHLGGKAFSHCSHL 186
519 KADENTPGE--GSAIAGVADSGSATRMOLSKLVTSLPWSALLTNHLKSTSEFPYYVL 576
187 ROHERSHNEKEKPHGCHLGGKAFTESSYLKRHERHTG-----EKPYEC 229
577 EPLGASPESTSKLOLVETKIDROGAANAFASTASCAPPTSAPAPSSASGPNOCVLCRLV 636
230 HVCGKAFTESSDLR-----HERHTGKRPYECCHLGGKAF 264
637 SCPRALRLHYGOHGGHPRPKCYGKAFSTRGNLRAHFGVHKHTSPAARQNSCPICOKAF 696
265 NHSSVLRREHRTHTGKRPYECNICGKAFNSYNNRLHRRVH---TGKRPYECPLCGKAF 320
697 TNAVTLQOHVRMH 709
```

```

Db 321 SKFENLRQHERHT 333
RESULT 27
US-09-864-761-38000
Sequence 38000, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEOTIC ACID PROBES USEFUL FO
FILE REFERENCE: Aemolica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 38000
LENGTH: 983
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL031277.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.98
OTHER INFORMATION: EST_HUMAN HIT: AU140221.1, EVALUE 1.00e-111
OTHER INFORMATION: SWISSPROT HIT: Q61116, EVALUE 1.00e-09
US-09-864-761-38000

Query Match 4.4%; Score 233.5; DB 10; Length 983;
```

Best Local Similarity 19.1%; Pred. No. 0.00023;
Matches 187; Conservative 100; Mismatches 383; Indels 307; Gaps 35;

OY 1 MAQETSSSLGFCG-EPARGGDASEHHPQVCAKCAQGFDPPEFLAHQNCCTDPP 59
Db 109 MSKEMTASSAFSSVYCMQDLSSGVKQK--AEGTKTPVQMSVLDSVHKKCSQSE- 165
OY 60 VVVIIIGQENPNSN-----SASAPRPGHSRSQVMDTEHS-----NPPDSSGSP-- 105
Db 166 -----GKEFESHVQPTCSAVKKRKPPTCMLOKVLNNEYNGIDLVENAD-CTRSPSP 219
OY 106 -----PDPPTPERRGEESGQFLVAATGAAGGGGIIIASPKIGATPLPP--EST 155
Db 220 CKSLEAOPDDELGP-----GSGFPAPVTESTPDVCPSSPALQTPSLSSQGLPLLPD 273
OY 156 PAPPPEPP-----PPRPGVSGHNLPLILEELRVLDGRQIHQNMOTDQICQVLLG 209
Db 274 PSSPPCPPLVATPPPPPLPLTVPLPAP----- 302
OY 210 SLGQTVGAPASPELP---GTGAASSTRPL-----PLFSPIKP-----AQTGKTAS 254
Db 303 -----SSASAPHCPRPLSNATASPLPLISPIVSPSPSPPIRPERLMAASGPPTLS 356
OY 255 SSSSSSSSGAEPKQAFPHLYHPLGSHQHPFSYGVGRSHKPPRAPSPALP--GSTDQLI 311
Db 357 SSSSSSSSSS-----SFSSSSSSSSPSPPLSAISSVSSGDNLE 396
OY 312 AS-PHIAFPPTTGLLAQCLGAARGLAASPGILKPKNGSGELGGEVYSSLEKPGRH 370
Db 397 ASIPMTSEK-----QEELNEG-LKPREPOSAAEDVVQV--ETFNKNF 438
OY 371 KCFKCAKVFEGSDSALOIHLSHTGERPYKCNVCGNFTTRGNLKVH----- 416
Db 439 VCVNCSPLFSLINDLTHLSHAEMPRKERCVOLEFKDLDSEIRFLHAGNITVCS 498
OY 417 -----FHRH-DEKYPHVQNMH-----PVREHLDYV----- 441
Db 499 VCKKEAFILCNLOHODLHPDKVCTHHEFESGTLRPQNTDPSKAHVEHQSLPEDPLE 558
OY 442 -----TSSGLP-----YGSVPERKEELEGTGCGG 468
Db 559 TSKEEBELNDSSELYTTIKTMASGIKTKRDPVRLGLNOHYPSEKPPFOYHNHRPMGIG 618
OY 469 VERKPLVASTTALATESLTLSTGTSTAVAPGL-----PTFNKFLVM 511
Db 619 VTAFTNFTTHNIPOFTTAIRCTKCGKGVDMPELKHILLACASASDKKRTTPKKNPPLK 678
OY 512 KAVEPRSKADENTPPGSEBSAIAVADSGSATRMOLSKLVTSLPNAALLTNHLKSTGSPF 571
Db 679 QTVQPK-----NGVVVDNDSGNKAFRRMGOPKRLNFSVELSKMSSNKLKLNALKKKNQLV 733
OY 572 FRYVLEPLGASPEFTSKIQQVLE-----KIDRQAVANASTASGAPTT 614
Db 734 QKAILQK-NKSAKOKADILKNACESSHICPCNREFTYIGSLKHAAFSCPKPLSPKK 792
OY 615 -----SAPAPSSASGPN-----OCVICRLVLSCPRALRLHYQHOGGERPFKC 657
Db 793 KYVSHSKKGGHSSPASSDKNSNHRRTADAELIKQSMQTP---LGKTRASSSGPTQV 848
OY 658 KYCGRAFSTRGNLR-AHFVGHKTSBARAONSCPIQCKKFTNA----- 699
Db 849 PLPSSSFRSKQNVKFAFASVSKKPKSSSSILRNSSPIRMAKITHVEGKKPKAVAKNHSQALS 908
OY 700 --VTLOOHVMMHILGQGLPNGSALSSEGGGAQENSEQSTAGSPGSPFOQSOQSPSEEE 757
Db 909 SKTSRSLHVRVQSKAVALQSKSTLASKKRTDRFNKTSRERSGGPVY---RSLQLAAD 964
OY 758 MSEEDEEDEEDVDVT 774
Db 965 LSENKREDSAKQELKD 981
RESULT 28

US-09-764-864-893
; Sequence 893, Application US/09764864
; Patent No. US20020132753A1

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 893
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-893

Query Match 4.4%; Score 232; DB 10; Length 361;
Best Local Similarity 20.8%; Pred. No. 9.6e-05;
Matches 73; Conservative 33; Mismatches 115; Indels 130; Gaps 8;

OY 364 EKPGRHKCFKCAVFGSDSALOIHLSHTGERPYKCNVCGNFTTRGNLKVHHRHEK 423
Db 121 EKP---HKTECGKSFNEKSTLIVHQRTHTGKPYECDVCGKFTQKSNLGVHQRTHSGE 177
OY 424 YPHQNMHPYREHLDYVITSSGLPYGMSVPEPEKAEAGTPEGGVKRLVASTLSA 483
Db 178 KPEECNECEKAFSOKSYTML-----HQKHTG---EKPEECNECEKAF 217
OY 484 TESTLTLSTGTSTAVAPGLPTFNKFLVMKAVEPRSKADENTPPGSEBSAIAVADSGSAT 543
Db 218 SQKSYLL-----IHQRHTEK----- 234
OY 544 RMOLSKLVTSLPNAALLTNHLKSTGSPFPYVLEPLGASPEFTSKL---QOLYEKIDROG 600
Db 235 -----PYKCNCEKAFREKSKLTIHQRI----- 257
OY 601 AVAVASTASGAPTSAPAPSSASGNOCVICRLVLSCPRALRLHYQHOGGERPFCKVC 660
Db 258 -----HTEKPYECPVCWKAFSQKSLIIRHQRTHTEGKPYACTEC 297
OY 661 GRAFSTRGNLRAHFVGHKTSBARAONSCPIQCKKFTNAVTLOOHVMMHILG 711
Db 298 GKAFREKST---FTVHQRTHTGKPYKTECGKAFQKSNLIVHQRTHNG 344

RESULT 29
US-09-764-864-954
; Sequence 954, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 954
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (328)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-954

Query Match 4.4%; Score 231; DB 10; Length 708;
Best Local Similarity 19.2%; Pred. No. 0.00021;
Matches 181; Conservative 88; Mismatches 328; Indels 348; Gaps 39;

```
OY 108 PTWGBRGESESSGQFLVATGTAAGGGGLTLASPKLGATPLPESTPAPPPPPPP 167
| | | | | : | : | : | | | | |
Db 13 PLMHGRCGTSSAAMNASSEGESFAGS---VOIP--GGTIVVELTP----- 55
| | | | | : | : | : | | | | |
OY 168 PGVSGHLNPLLEELRVLQORQIOMOMTEQICROVLLGSLGQTVGAPASPELPGT 227
| | | | | : | : | : | | | | |
Db 56 -----DINHC---GICKQ--QFNLDARVAHKGQGLTGT 86
| | | | | : | : | : | | | | |
OY 228 GAASSTKPLPLPSPKPAQOTGKTASSSSSSSSSSGAEPKQAFH---LYHPLGSH-- 282
| | | | | : | : | : | | | | |
Db 87 SAAPSTVOFVSEETVPATOTOTTTITITSETQITIVSABEFVEHGOYTLPTESNMO 146
| | | | | : | : | : | | | | |
OY 283 ---PVSQVGVGSHKPTPAPSPALPGSTDLIASPHLAPGTGTLAAQCGAARGLEAA 339
| | | | | : | : | : | | | | |
Db 147 TATVLSLAKSTKTKPTTPAQ-----KRLNCCYPCGQ----- 179
| | | | | : | : | : | | | | |
OY 340 ASPGLLKPRNGSGELGCGVVISLEKPGGNHRCFCAKVFSGDSALQIHLRSHTEGPEYK 399
| | | | | : | : | : | | | | |
Db 180 ----FKTAYGKMDMRHLKIHGDKP---HKCEVCGKCFSRKDKLTHMRCHTGVKPEYK 231
| | | | | : | : | : | | | | |
OY 400 CNVGNKRTTRGNLKVHPIHREKTP--HYQMNHPVPEHLDYVITSSGLPYGMSVPEKA 458
| | | | | : | : | : | | | | |
Db 232 CKTCYAAADSSSLKHLRIHSDERPFKCQICPY----- 265
| | | | | : | : | : | | | | |
OY 459 EEEACTPGGVERKPLVASTTALATESLTLLSTGTAVAPGLPTFN-----KFLMK 512
| | | | | : | : | : | | | | |
Db 266 -----ASRNSQLTVHLR-----SHTGAP-----FOCWLSAKFKLSS 299
| | | | | : | : | : | | | | |
OY 513 AVEPKSKADENTPPGSEGSALAGVADSGSATRMQLSKLVTLSPWALLTNHLK---STGS 569
| | | | | : | : | : | | | | |
Db 300 DLKRMHRYHSGEKP-----FKCEFCNVRCITMK--GXLSKSHIRIKHSGNN 341
| | | | | : | : | : | | | | |
OY 570 FFPFVLEPLGASPSSTSKLQOLVEKIDROGAVALAVASTASAPITTSAPABSSSASGPNOC 629
| | | | | : | : | : | | | | |
Db 342 FKCPH-CDFLGOSKALRLRHRSRV-----HOSEHPKGC 372
| | | | | : | : | : | | | | |
OY 630 VICTLVLSCPRALRLHYGONGGERPFCKVCGRAFSR--GNLRANF----- 674
| | | | | : | : | : | | | | |
Db 373 SFGSVSSCSKALRIHERIHCTRPFCNVC--SFDTKQPSNLSKHKMKFHDIMKTEAL 430
| | | | | : | : | : | | | | |
OY 675 ---VGHKTSPA---ARAQNSCPICQCKFTNAVTLQOHVHMLGQIIPNGSALSSEG 725
| | | | | : | : | : | | | | |
Db 431 ERKDIQRQSRQVAKLDKKSFFHCIDICDASFMREDSLRSHKRON-----SEYSEK 481
| | | | | : | : | : | | | | |
OY 726 GA-----AQENSSFOBTASGSGSPPOQSPREEMSEEEDEEEDVTDDESL 778
| | | | | : | : | : | | | | |
Db 482 NSDVTVLOFQIDPSKQAPATPLTVGHLOVP--LOPSOVPOFSE----- 521
| | | | | : | : | : | | | | |
OY 779 AGRGESEGEKAISVGDSEEVSGAEDEVATSVAPTTV-----KEMDSNEKAP-- 827
| | | | | : | : | : | | | | |
Db 522 -----GRVITIV---GHQVQANTIVQAAAAAANINVPALVAQNEELPGNSRLQIL 570
| | | | | : | : | : | | | | |
OY 828 -QHTLPPRPDPNDLHPORMEQGSDVSGAMEBEAKLEGISSPMALTOBEGESTPLV 886
| | | | | : | : | : | | | | |
Db 571 RQVSLIAPROP-----SRCESEAG-----ANTQPAVLLTTH-EQIDGATL 609
| | | | | : | : | : | | | | |
OY 887 EELNLPKAKKPDGESSGRKACEVCGOSFPOTALEHOKTHPKDGLFTC-----VF 939
| | | | | : | : | : | | | | |
Db 610 HQTLLPTA-SGGQEGSG-----NOTFTTSSGI-----TCDFEGLNAL 647
| | | | | : | : | : | | | | |
OY 940 CROGFLDRATLKKHMLLHNVPPFAPHGPONITLSLPPCCSS 984
| | | | | : | : | : | | | | |
Db 648 IOEGTAETVVS-----DGGONIAVATTAPVFFSS 677
| | | | | : | : | : | | | | |
```

```
RESULT 30
US-09-734-329-2
; Sequence 2, Application US/09734329
; Patent No. US20020156031A1
; GENERAL INFORMATION:
; APPLICANT: de CROMBRUGHE, BENOIT
; APPLICANT: NAKASHIMA, KAZUHIKA
```

```
; APPLICANT: ZHOU, XIN
; TITLE OF INVENTION: MASTER BONE FORMATION TRANSCRIPTION FACTOR:
; TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: UTXC:666
; CURRENT APPLICATION NUMBER: US/09/734,329
; CURRENT FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-734-329-2

Query Match 4.3%; Score 229; DB 9; Length 428;
Best Local Similarity 24.0%; Pred. No. 0.00016;
Matches 111; Conservative 26; Mismatches 137; Indels 186; Gaps 20;

OY 130 TAAGGGG---LILASPKL--GATPLPESTPAPPPPPPPPGVSGHLNPLILEE 183
| | | | | : | : | : | | | | |
Db 37 TITLKGSTKKRYADLSAPKTMGDATPAPFSSINGLSPAGSPAP--ASGYAN----- 87
| | | | | : | : | : | | | | |
OY 184 LRVLQORQIOMOMTEQICROVLLGSLGQTVGAPASPELPGTGAASSTKPLPLPSPI 243
| | | | | : | : | : | | | | |
Db 88 -----DYPFPFSPFGPTGAODPGLLV----- 110
| | | | | : | : | : | | | | |
OY 244 KPAOTGKTASSSSSSSSGAEPKQAFHLYHPLGSHQHFV--GGV--GRSHKPTP-- 297
| | | | | : | : | : | | | | |
Db 111 -----KSHSSDCLPSVYTSLDMTHPYGSWKAGIHAGISPGGNTPTPMD 157
| | | | | : | : | : | | | | |
OY 298 -----APSPALPGSTDL-----IASPHLAFPGT 321
| | | | | : | : | : | | | | |
Db 158 MHPRGNVLGGGGGDDGLGTLSTGAPQPLNPLQPLTYSDFAPLNPARYPAHLLQPCP 217
| | | | | : | : | : | | | | |
OY 322 TGLL-----AAQCGAARGLEAASPGLKPKNGSGELGYEYIS----- 362
| | | | | : | : | : | | | | |
Db 218 QHVLPODVYKPKAVNSGQLGGS---GAAKPPRGAGTGSGGAGSAGRSTDCPCNOC 274
| | | | | : | : | : | | | | |
OY 363 LEKPGGR-----HKCRF--CAKVFSGDSALQIHLRSHTEGPEYKCN--VCGNRETT 409
| | | | | : | : | : | | | | |
Db 275 LERLGAANAAGLRKKRPHSCHIPGCGYVGKASHLKAHLMWHTGERPFVCMWLFCKGRFTR 334
| | | | | : | : | : | | | | |
OY 410 RGNLVHFRH-REK-----YPHVOMNHPVP----- 435
| | | | | : | : | : | | | | |
Db 335 SDELEHVRHTTRREKFKTCLCSKRFTRSDHLSKHQRTGEPGPGPPSPKELGEGRSV 394
| | | | | : | : | : | | | | |
OY 436 --EHLDVYTSSGLPYGMSVPEKAEEEAGTGGGVERRPLV 475
| | | | | : | : | : | | | | |
Db 395 GEEBANOPPRRSTSP---APPEKAH-----GGSPEQGNLL 426
| | | | | : | : | : | | | | |

RESULT 31
US-09-864-761-33653
; Sequence 33653, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
```

```

1  PRIOR APPLICATION NUMBER: US 60/236,359
2  PRIOR FILING DATE: 2000-09-27
3  PRIOR APPLICATION NUMBER: PCT/US01/00666
4  PRIOR FILING DATE: 2001-01-30
5  PRIOR APPLICATION NUMBER: PCT/US01/00667
6  PRIOR FILING DATE: 2001-01-30
7  PRIOR APPLICATION NUMBER: PCT/US01/00664
8  PRIOR FILING DATE: 2001-01-30
9  PRIOR APPLICATION NUMBER: PCT/US01/00669
10 PRIOR FILING DATE: 2001-01-30
11 PRIOR APPLICATION NUMBER: PCT/US01/00665
12 PRIOR FILING DATE: 2001-01-30
13 PRIOR APPLICATION NUMBER: PCT/US01/00668
14 PRIOR FILING DATE: 2001-01-30
15 PRIOR APPLICATION NUMBER: PCT/US01/00663
16 PRIOR FILING DATE: 2001-01-30
17 PRIOR APPLICATION NUMBER: PCT/US01/00662
18 PRIOR FILING DATE: 2001-01-30
19 PRIOR APPLICATION NUMBER: PCT/US01/00661
20 PRIOR FILING DATE: 2001-01-30
21 PRIOR APPLICATION NUMBER: PCT/US01/00670
22 PRIOR FILING DATE: 2001-01-30
23 PRIOR APPLICATION NUMBER: US 60/234,687
24 PRIOR FILING DATE: 2000-09-21
25 PRIOR APPLICATION NUMBER: US 09/608,408
26 PRIOR FILING DATE: 2000-06-30
27 PRIOR APPLICATION NUMBER: US 09/774,203
28 PRIOR FILING DATE: 2001-01-29
29 NUMBER OF SEQ ID NOS: 49117
30 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
31 SEQ ID NO 33653
32 LENGTH: 525
33 TYPE: PRT
34 ORGANISM: Homo sapiens
35 FEATURE:
36 OTHER INFORMATION: MAP TO 284476.6
37 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.8
38 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
39 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
40 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
41 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
42 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
43 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
44 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
45 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
46 OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALU0 1.00e-106
47 OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALU0 1.00e-106
48 US-09-864-761-33653
49
50 Query Match 4.3% Score 227.5; DB 10; Length 525;
51 Best Local Similarity 22.5%; Pred. No. 0.00023;
52 Matches 84; Conservative 37; Mismatches 134; Indels 119; Gaps 10;
53
54 364 EKGGRKRCFCACAVFGSDSALQIHLSRSTGERPYKCNVCNFGNFTTRGNLKVHFRHREK 423
55 187 EKP--HECDGCGAFKTRNRLCMHQLIHTGKRPYKCNCCGKAFOFKHSITLHGRHTGE 243
56 424 YPVMQNPHPVPEHLDYVITSSGLPYGMSVPEKAEELAGTPCGGVGRKPLVASTTALSA 483
57 244 KPL-----ECEECC-----KAFSG 257
58 484 TESLTLLSTGTSTAVABGLPTFNKFLMKAVEPKSKADETTPGSEGSALAGVADSGSAT 543
59 258 SSDLT-----KHIRLHTGRRPYGSKC-----GRATSRSDLSNKK 293
60 544 RMOLSKLVYSLPSW-----ALLTNHLKSTGSPFPYVLEPIGASBSETSKIQOLYEKI 596
61 294 RIHTREKHVYCGPOCGKDFSIKAEILTKH-RIHTTEKRYKCEEGCGKAFRNCKRR----- 346
62 597 DNGGNVAVASTAGAPITTSAPAPSSASGPNOCVLCRLVSCPRALRLHNGOGERPEK 656
63 347 -----AHREHTGKRPYQCCROCGKFPDCKHCLTLHRIHTGKERPYK 387

```

```

QY 657 CXYCGRAFTSTGNLRANFVGHKTSAPAAQNSCPICQKFTNATVLOOHVAMHLG----- 711
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 388 CLRCGKAIFSGKSNL-----TNHRIHTGKPRHKCEVCGMARHSSVLRLOKRIHTGKEPYT 443
QY 712 ----GQIPNGGSAL 721
      |:::|:::|
Db 444 CSECGTSTFRQGSAL 457

RESULT 32
US-09-764-864-956
; Sequence 956, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 956
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (315)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (340)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-956

Query Match
Best Local Similarity 22.9%; Score 226; DB 10; Length 378;
Matches 96; Conservative 27; Mismatches 159; Indels 138; Gaps 13;

QY 621 SSASGPNOCVICTLVLSCPRALRLHYGONGGERPEKCKVCGRAFTSTGNLRANFVGHKTS 680
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 10 SHREGVFKCPLCSKVFPPSSLDQHLGDHSSSESHFLCVSCGLAFTGTEALLAHNRATPN 69
QY 681 PARAAQNSCPICQKFTNATVLOOHVAMHLGQIPNGGSALSSGGGAQOQNSSQSTASG 740
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 70 PL-----HSCP-CGKTFNLTFLYHRRH-----G 94
QY 741 PGSFPOPOSOOPSEEE-----EMSEEEDEDEEEDVTDSDIAGRS----- 783
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 95 VGVVPLPTTVLP-PEEYVIGFPERPATGDEADPEPRPYSEESISAGARAGTYRCLCSR 153
QY 784 ESGGEKAISVRGDSSEYSGAEDEVATVAAPTTVKEMDSNEKARQHTLPPPPRPDNLDA 843
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 154 EFG--KALQLTRHQRFHRLERRHKCSICGKMKKSHVNRNLRTHTGERRFPQPD--- 207
QY 844 PQMEQGTSTVSGAMEBEAKLEGISPPAALQLQEGEGSTPLVEELNLPRAMKKDPGESS 903
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 208 -----CSKPRNSPANTLANRHLTHGERRP 230
QY 904 GRKACEVCGSFPQTALAEHQKTHPKDQPL-----FT 936
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 231 YR--CGDDCKAFTQSSSTLRQRLVHAQHFPRYKQSCGVAFNRPYRLIMNRHYNTTGEPYK 288
QY 937 CYVFCRGFLDRALILKKMLLANHVOVPPRPARIGPQNI-----ATLSLVPG---CSSSTPSP 988
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 289 CRECPRSFTLRRLLEVLVHLYVH-----AGRNHPLRLIMQCLPLLTAPGARPLSCRCXQAP 343

```

```

; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42897
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007228.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P10078, EVALU0 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AL042491.2, EVALU0 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AL042491.2, EVALU0 0.00e+00
US-09-864-761-42897

Query Match          4.3%; Score 226; DB 10; Length 492;
Best Local Similarity 22.9%; Pred. No. 0 00027;
Matches 90; Conservative 32; Mismatches 131; Indels 140; Gaps 14;

QY 371 KCRFCAKVGSDSALQIHLRSHTGERPKYKNCVGNRPTTNGNLKVHFRHREKYP----- 425
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 121 KCNECKKFTQSSSLVHQRIHTGEKPKYKNECGKAFSDSSGAFARHQCHGCKKPYECIE 180
-----HVOGNPH-----PYPEHLDYIT 443
QY 426 -----HVOGNPH-----PYPEHLDYIT 443
Db 181 CGKAFIQTSLIRHWRYHTGEKPFDCIDCGKAFSDHIGLGNHRRIRHTGEKPKYKCD--VC 238
```

```

QY 444 SSGLPYGMSVPEKAEDEAGTP--GGVERKPLVASTALSTESLTJLSTGTSTAVAPGL 502
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 239 HKSFRRGSSSLTVHQRIHTGEKPYECDCRK-----AFSHASLT-----QHQRVHSGE 286
QY 503 PTEKFKYLMKAVEPKSADENTPPGSEGSALAGVADSSATRTMQLSKLVTSLPSMALTLN 562
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 287 KRP-----KCR-----ECGKAFRONIH-----LAS 306
QY 563 HLK--STGSFPFVLEPLDGLASPSSETSKL---QOLVEKIDRGAVAVASTAGAPTTSAPA 618
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 307 HLRIHTGEKPE--CAGCGSFSISSQLATHORI----- 338
QY 619 PSSASGPNQCVICLRVLSGPRALRLHYGHGGERPKCYKAGRAFSTRGNLRAHFVGHK 678
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 339 --HTGEKPYECKVCKSAFTQKLAHLAQHOKHTHGEKPYECKGKAFSQ----TTHLIQHQ 392
QY 679 TSPAARAQNSCPICQKKFTNAVTIQQHVRMLG 711
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 393 RVHTGEKPKYKMECGKAFGDNSSCTQHRLHTG 425

RESULT 34
US-09-974-298-118
; Sequence 118, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 118
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 5665139CD1
US-09-974-298-118

Query Match          4.3%; Score 226; DB 9; Length 516;
Best Local Similarity 24.3%; Pred. No. 0.00027;
Matches 87; Conservative 32; Mismatches 133; Indels 106; Gaps 9;

QY 359 VSSLEKPGGRHKRCFCAKVPDSDALQIHLRSHTGERPKYKNCVGNRPTTNGNLKVHFR 418
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 253 IHTGEKPR---YRCNVCGKYVHHHSHLQHQRIHTGEKPKYKNECGKAFVSHKSSLVNHW 309
QY 419 RHREKYPHVOGNPHRPVPEHLDYITSSGLPYGMSVPEKAEDEAGTPGGVERKPLVAST 478
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 310 IHNGEKPKYKNECGKAFVSHKSSLVNHWRIHTGEK--PYKCNE---CGKVFSSNSYLA-- 361
QY 479 TALSATSLTLTGTSTAAVAPGLPTFNKFKYLMKAVEPKSKADENTPPSEGSALINGVAD 538
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 362 -----OHLTIHAGEKPKYKDECCAKASONS----- 386
QY 539 SGSATRMQLSKLVTSLPSMALTLNLKSTGSFPFVLEPLDGLASPSSETSKLQ-----QLV 593
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 387 -----HLVQNHRIHTGE--KPYKDECGKAFVPSQNSYLAVHWRHTHG 425
QY 594 EKIDRGAVAVASTAGAPTTSAPASSASGPNQCVICLRVLSGPRALRLHYGHGGER 653
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 426 EK-----AYKNECGKAFGLNSLAHHRKIHTGEK 455
QY 654 PFKCVGGRATSTRGNLRAHFVGHKTSPPAARAQNSCPICQKKFTNAVTIQQHVRMLG 711
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 456 PFKNECGKAFSSMSSLTNHAH--TGEKHKFKNECGKLFKRNDSYLVLRHORFHAG 509
```



```
Db 48 FLGAAGAPEGSGSSSSSSSGGGGGG-----GGSN-----SSSSSTFN 87
Oy 369 LRSHGGERPKVCNRRFTTRGNLKVHFRHREKYPHVQNNPHVPEHLDYITSSGLP 448
Db 88 POADTGEQPYE-----HLTAESFPDISLNNEKVLVETSPSQTTRLP 129
Oy 449 ---YGMVSPPEKAEAEAGT-----PGGVERKPLVASTTALSTATSESL 487
Db 130 PITYTGRSLEAPRNSGNTLWPEPLFSLVSGVSMTPNPASSSSAPSPAASSA-SASOSP 188
Oy 488 TLLSTGTST-----AVAPGLPTFNKFLVLMKAVEPKSKADENTPPGSGSAGIAGVADSGS 541
Db 189 PLSCAVPSNDSSPIYSAAPTEPPTNDIF---PEPQSAF---PGSAGTALQYPPPAYP 241
Oy 542 ATRMOLSKLVTSIPSWALLTNHLKSTGSPFPYVLEPLGASPSETSKLQOLVEKIDROGA 601
Db 242 AAKGFGQ--VPMIPDYL-----FPQOGDLGLGTPDOKPFOGL-BSRTQPS 285
Oy 602 VAVASTAGAPTTTAP-----APSSASGPNQ-----CVI--C 632
Db 286 LTPLTSTIKAFATQSGSODLKALNTSYOSOLIKPSRMKKYPNRPESKTPRHERPYACPVESC 345
Oy 633 LRVLSCPRALRLHYGQHGGERPKVCYCGRAFSTRGNLRAHFVGHKTSPARAQNSCPIC 692
Db 346 DRRFSRSDLTFRHRIHRTGQKPCRCICMNFSSDHLTTHIRTH---TGKPPFADIC 401
Oy 693 OKKFTNAVTLQOHVHNLGGQIPNGSALSSEGGAQENSSEOSTAGSPGSPPOPOSOP 752
Db 402 GRKFARSDERKRRTKIH-----LRQDKKADKSVASSATSSLSYSPS-----445
Oy 753 SPEEEMSEEEDEEEDVTDSDSLAGRSGESGKAISVRDSEEVSGAEEVATSYA 812
Db 446 -----VATSTP 451
Oy 813 APTTVKEMDSNEKAPQHTLPPPPP-----PDNLDPQPMEOGTSDVSGAMEEAKLEGI 867
Db 452 SFTV-----TSYSPATTSYSPVPFTSFSSPSSSTYSPVH-----SGF 490
Oy 868 SSPMAALTQEGEGTSPVLEEINLPEAMKKDQEGSSGRKACEVCGSPPTOTALEEHOKT 927
Db 491 PPSVATTYS-----SVPPAF---PAOVSSFPSSAVT-NSFSASTGLSDMTAT 534

RESULT 38
US-09-902-941-337
; Sequence 337, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshitiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Mard W.
; APPLICANT: Manerakis, Margarita
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvik, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902.941
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 337
; LENGTH: 543
; TYPE: prt
; ORGANISM: Homo sapiens
; US-09-902-941-337
```

Query Match 4.2%; Score 224.5; DB 9; Length 543;

```
Best Local Similarity 20.6%; Pred. No. 0.00033;
Matches 148; Conservative 67; Mismatches 254; Indels 251; Gaps 28;

Oy 269 QAFPHLYHPLGQHPFSVGVGRSHKPRPAEPAALPGSDOLASPHLAFPGTGLLAQ 328
Db 5 KAEMQLMSPLIDISDFG---SFPSPMDNYPKL---BEMMLLS-----NG-ApQ 47
Oy 329 CIGARGLAEAAAPGLKPKNSGELGYEIVLSLEKPGGRHKRCFAKFGSDSALOIH 388
Db 48 FLGAAGAPEGSGSSSSSSSGGGGGG-----GGSN-----SSSSSTFN 87
Oy 389 LRSHGGERPKVCNRRFTTRGNLKVHFRHREKYPHVQNNPHVPEHLDYITSSGLP 448
Db 88 POADTGEQPYE-----HLTAESFPDISLNNEKVLVETSPSQTTRLP 129
Oy 449 ---YGMVSPPEKAEAEAGT-----PGGVERKPLVASTTALSTATSESL 487
Db 130 PITYTGRSLEAPRNSGNTLWPEPLFSLVSGVSMTPNPASSSSAPSPAASSA-SASOSP 188
Oy 488 TLLSTGTST-----AVAPGLPTFNKFLVLMKAVEPKSKADENTPPGSGSAGIAGVADSGS 541
Db 189 PLSCAVPSNDSSPIYSAAPTEPPTNDIF---PEPQSAF---PGSAGTALQYPPPAYP 241
Oy 542 ATRMOLSKLVTSIPSWALLTNHLKSTGSPFPYVLEPLGASPSETSKLQOLVEKIDROGA 601
Db 242 AAKGFGQ--VPMIPDYL-----FPQOGDLGLGTPDOKPFOGL-BSRTQPS 285
Oy 602 VAVASTAGAPTTTAP-----APSSASGPNQ-----CVI--C 632
Db 286 LTPLTSTIKAFATQSGSODLKALNTSYOSOLIKPSRMKKYPNRPESKTPRHERPYACPVESC 345
Oy 633 LRVLSCPRALRLHYGQHGGERPKVCYCGRAFSTRGNLRAHFVGHKTSPARAQNSCPIC 692
Db 346 DRRFSRSDLTFRHRIHRTGQKPCRCICMNFSSDHLTTHIRTH---TGKPPFADIC 401
Oy 693 OKKFTNAVTLQOHVHNLGGQIPNGSALSSEGGAQENSSEOSTAGSPGSPPOPOSOP 752
Db 402 GRKFARSDERKRRTKIH-----LRQDKKADKSVASSATSSLSYSPS-----445
Oy 753 SPEEEMSEEEDEEEDVTDSDSLAGRSGESGKAISVRDSEEVSGAEEVATSYA 812
Db 446 -----VATSTP 451
Oy 813 APTTVKEMDSNEKAPQHTLPPPPP-----PDNLDPQPMEOGTSDVSGAMEEAKLEGI 867
Db 452 SFTV-----TSYSPATTSYSPVPFTSFSSPSSSTYSPVH-----SGF 490
Oy 868 SSPMAALTQEGEGTSPVLEEINLPEAMKKDQEGSSGRKACEVCGSPPTOTALEEHOKT 927
Db 491 PPSVATTYS-----SVPPAF---PAOVSSFPSSAVT-NSFSASTGLSDMTAT 534

RESULT 39
US-09-849-626-337
; Sequence 337, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Tonglong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849.626
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 337
; LENGTH: 543
```



```

"      TYPE: PRT
";      ORGANISM: Homo sapiens
JS-09-849-626-337

```

Query Match	4.28;	Score 224.5;	DB 9;	Length 543;
Best Local Similarity	20.68;	Pred. No. 0.00033;		
Matches 148; Conservative	67;	Mismatches 254;	Indels 251;	Gaps 28;

```

RESULT 40
US-09-929-315-4
; Sequence 4, Application US/09929315
; Patent NO. US20020082394A1
; GENERAL INFORMATION:
; APPLICANT: Call, Katherine M.
; APPLICANT: Glaser, Thomas M.
; APPLICANT: Ito, Caryn Y.
; APPLICANT: Buckler, Alan J.
; APPLICANT: Pelletier, Jerry
; APPLICANT: Haber, Daniel A.
; APPLICANT: Rose, David A.
; APPLICANT: Houseman, David E.
; APPLICANT: Bruening, Wendy
; APPLICANT: Darveau, Andre
; TITLE OF INVENTION: Localization and Characterization of the

```

```

1 TITLE OF INVENTION: Wlms/'Tumor Gene
2 FILE REFERENCE: 0050.1312-013
3 CURRENT APPLICATION NUMBER: US/09/929,315
4 CURRENT FILING DATE: 2001-08-14
5 PRIOR APPLICATION NUMBER: US 09/037,179
6 PRIOR FILING DATE: 1998-03-09
7 PRIOR APPLICATION NUMBER: US 08/102,942
8 PRIOR FILING DATE: 1993-08-02
9 PRIOR APPLICATION NUMBER: US 07/614,161
10 PRIOR FILING DATE: 1990-11-13
11 PRIOR APPLICATION NUMBER: US 07/435,780
12 PRIOR FILING DATE: 1989-11-13
13 PRIOR APPLICATION NUMBER: US 07/795,323
14 PRIOR FILING DATE: 1991-11-20
15 NUMBER OF SEQ ID NOS: 21
16 SOFTWARE: FastSeq for Windows Version 4.0
17 SEQ ID NO 4
18 LENGTH: 449
19 TYPE: prt
20 ORGANISM: Unknown
21 FEATURE:
22 OTHER INFORMATION: Murine
23 US-09-929-315-4

```

Search completed: January 13, 2003, 15:23:42
Job time : 31.9851 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:17:40 : Search time 38.0568 Seconds
(without alignments)
3518.868 Million cell updates/sec

Title: US-09-988-117-1
Perfect score: 5277
Sequence: IMAHERSSRSRLGVPAGEPAE.....PSTSTGLSPFPKDDPTIP 1005

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : A_Geneseq_101002:*

1: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
18: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
19: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
20: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
21: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
22: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
23: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5277	100.0	1005	23	AAE17954	Human Sal2 protein
2	5272	99.9	1005	23	AAE17967	Human Sal2 protein
3	5269	99.8	1005	23	AAE17968	Human Sal2 protein
4	5171	98.0	1007	22	AAW78838	Human protein SEQ
5	5161	97.8	1019	22	AAW78822	Human protein SEQ
6	4677.5	88.6	1002	23	AAE17955	Mouse Sal2 protein
7	877.5	16.6	813	22	AAE17955	Human protein sequ
8	784	14.9	1373	22	ABE62368	Drosophila melanog
9	733	13.9	172	22	AAO1962	Human polypeptide
10	726	13.8	1267	22	ABE61050	Drosophila melanog

11	596	11.3	200	22	AAE2796	Human protein sequ
12	489	9.3	330	22	ABE30894	Peptide #3545 enco
13	489	9.3	330	22	ABE36076	Peptide #3582 enco
14	489	9.3	330	22	AAE56858	Human brain expres
15	489	9.3	330	22	AAE69245	Human bone marrow
16	489	9.3	330	22	AAE17074	Peptide #3508 enco
17	489	9.3	330	22	AAE29568	Peptide #3605 enco
18	489	9.3	330	22	AAE04775	Peptide #3457 enco
19	489	9.3	330	22	ABE38855	Human peptide enco
20	441	8.4	336	22	ABE32272	Peptide #4923 enco
21	441	8.4	336	22	ABE37530	Peptide #5036 enco
22	441	8.4	336	22	AAE58186	Human brain expres
23	441	8.4	336	22	AAE70641	Human bone marrow
24	441	8.4	336	22	AAE18485	Peptide #4993 enco
25	441	8.4	336	22	AAE30956	Peptide #4757 enco
26	441	8.4	336	22	AAE06075	Human peptide enco
27	441	8.4	336	22	ABE40328	Human zinc finger
28	343	6.5	702	23	ABE04333	Human nucleic acid
29	341	6.5	799	21	AAE21003	Human protein SEQ
30	334	6.3	927	22	AAE80283	Human protein sequ
31	333.5	6.3	619	22	AAE95103	Human protein sequ
32	329.5	6.2	498	22	AAE94669	Human protein sequ
33	324.5	6.1	725	22	AAE95116	Human protein SEQ
34	320.5	6.1	869	22	AAE79299	Novel human diagno
35	311.5	5.9	882	22	ABE06465	Novel human diagno
36	311.5	5.9	934	22	ABE69958	Drosophila melanog
37	308.5	5.8	1893	22	ABE59829	Drosophila melanog
38	303	5.7	722	22	AAE39029	Human polypeptide
39	298.5	5.7	744	22	ABE59449	Drosophila melanog
40	298.5	5.7	744	22	ABE66002	Drosophila melanog
41	298.5	5.7	744	22	ABE66572	Drosophila melanog
42	295	5.6	496	22	AAE95788	Human protein sequ
43	293.5	5.6	1472	22	ABE00399	Novel human diagno
44	293	5.6	820	22	AAE131578	Drosophila melanog
45	293	5.6	820	22	ABE65565	Mouse combined DNA
46	293	5.6	884	22	AAE99334	Novel human diagno
47	291	5.5	1685	22	ABE62718	Novel human diagno
48	290.5	5.5	691	22	AAE94433	Human protein sequ
49	289.5	5.5	1185	20	AAE33497	Human atrophin I p
50	289	5.5	880	22	AAE99336	Mouse Riti alpha'
51	288	5.5	1220	22	AAE00831	Novel human secret
52	285.5	5.4	832	22	ABE10543	Novel human diagno
53	284.5	5.4	1196	22	ABE28743	Novel human diagno
54	284	5.4	582	23	ABE97429	Novel human protei
55	283.5	5.4	883	22	ABE22743	Novel human diagno
56	283.5	5.4	1252	22	AAE79739	Human protein SEQ
57	282	5.3	613	22	AAE95862	Human protein sequ
58	281.5	5.3	491	21	ABE52155	Human secreted pro
59	281.5	5.3	624	23	ABE57324	Mouse ischaemic co
60	281.5	5.3	894	22	AAE99335	Human Riti protein
61	280.5	5.3	719	22	ABE16953	Novel human diagno
62	279.5	5.3	973	22	ABE21022	Novel human diagno
63	278.5	5.3	890	22	AAE99339	Human Riti alpha'
64	278	5.3	755	22	AAE40916	Human polypeptide
65	277	5.2	1054	22	ABE57993	Drosophila melanog
66	276.5	5.2	742	23	ABE05703	Human nucleic acid
67	276.5	5.2	1342	22	ABE01726	Novel human diagno
68	275	5.2	809	22	AAE38689	Human polypeptide
69	275	5.2	2703	22	ABE63299	Drosophila melanog
70	274.5	5.2	730	22	AAE93278	Human protein sequ
71	273.5	5.2	356	22	ABE60575	Drosophila melanog
72	273	5.2	675	22	AAE94388	Human protein sequ
73	273	5.2	714	22	ABE17368	Novel human diagno
74	273	5.2	812	22	AAE99337	Mouse Riti beta pr
75	272.5	5.2	878	22	AAE29528	Novel human secret
76	271.5	5.1	803	22	AAE95278	Human protein sequ
77	271	5.1	823	22	AAE99340	Human Riti beta pr
78	270	5.1	839	22	AAE93280	Human protein sequ
79	269.5	5.1	688	23	ABE79480	Human zinc finger
80	269.5	5.1	720	22	AAE94329	Human protein sequ
81	269.5	5.1	1203	22	ABE61375	Drosophila melanog
82	269	5.1	751	22	AAE39508	Human polypeptide
83	268.5	5.1	841	22	ABE16167	Novel human diagno

84	268	5.1	567	22	AAM38658	Human polypeptide
85	268	5.1	751	22	AAM39130	Human polypeptide
86	268	5.1	1205	22	ABG05068	Novel human diago
87	268	5.1	1205	22	ABG15312	Novel human diago
88	268	5.1	1214	22	ABG27121	Novel human diago
89	267.5	5.1	515	22	ABB31309	Peptide #3960 enco
90	267.5	5.1	515	22	ABB36515	Peptide #4021 enco
91	267.5	5.1	515	22	ABB21854	Peptide #3853 enco
92	267.5	5.1	515	22	AAM57278	Human brain expres
93	267.5	5.1	515	22	AAM69683	Human bone marrow
94	267.5	5.1	515	22	AAM17493	Peptide #3929 enco
95	267.5	5.1	515	22	AAM30016	Peptide #4053 enco
96	267.5	5.1	515	22	AAM05170	Peptide #3852 enco
97	267.5	5.1	515	23	ABG39299	Human peptide enco
98	266.5	5.1	2639	22	ABG15016	Novel human diago
99	266	5.0	577	22	AAM15930	Human novel secret
100	266	5.0	645	22	ABG28285	Novel human diago

ALIGNMENTS

RESULT 1

ID AAE17954 standard; protein; 1005 AA.

XX AAE17954;

DT 07-MAY-2002 (first entry)

XX Human Sal2 protein.

KW Human; proliferative disorder; tumor host range mutant virus; cancer;
T-HR mutant; Sal2 protein; ovarian tumour; chromosome 14q12.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 15

FT /note= "Encoded by TGC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 22..23

FT /note= "Encoded by bases 4611-5022 of the inverse complementary strand of the sequence shown as

FT Misc-difference 120

FT /note= "Encoded by CCA of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 545

FT /note= "Encoded by CGC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 552

FT /note= "Encoded by GTG of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 573

FT /note= "Encoded by TTC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 575

FT /note= "Encoded by TAT of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 576

FT /note= "Encoded by GTG of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 577

FT /note= "Encoded by ATC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 578

FT /note= "Encoded by GAG of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 579

FT /note= "Encoded by CCC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

XX

PN	NO200204596-A2.
XX	
XX	17-JAN-2002.
PD	
XX	05-JUL-2001; 2001WO-US21354.
PF	
XX	07-JUL-2000; 2000US-216723P.
PR	19-MAR-2001; 2001US-0812471.
PR	19-MAR-2001; 2001US-0812633.
XX	(HARD) HARVARD COLLEGE.
PA	(BGHM) BRIGHAM & WOMENS HOSPITAL INC.
PI	Benjamin TL, Li D, Mok SC, Cramer DW, Ma Y;
XX	WPI; 2002-164637/21.
DR	N-PSDB; AAD28651.
DR	
XX	

detecting protein involved in susceptibility to proliferative disease, by infecting normal and abnormal proliferating cells with mutant virus, detecting mutated protein allowing growth of mutant on abnormal cells

Disclosure; Page 77-79; 92pp; English.

CC The invention relates to a method for the identification of genes
CC and their encoded proteins involved in susceptibility to proliferative
CC disorders, including cancer using a tumor host range mutant (T-HR
CC mutant) virus. The invention also provides the use of Sal2 genes and
CC proteins in methods of identifying a mammal having, or at a risk of
CC acquiring a proliferative disease. T-HR mutants are used to kill cancer
CC cells such as one carrying a Sal2 alteration. Transgenic and knockout
CC mouse comprising Sal2 nucleic acid are useful as research tools to
CC determine genetic and physiological features of cancer and for
CC identifying compounds that can affect ovarian and other tumours. The
CC present sequence is human Sal2 protein. The Sal2 gene is located on
CC chromosome 14q12.

Sequence 1005 AA:

Query Match 100.0%; Score 5277; DB 23; Length 1005;
Best Local Similarity 100.0%; Pred. No. 4.2e-295;
Matches 1005; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	1	MAHESRSSRLGVPAAGPAETGSDASEDDHPOVCAKCCAQFDPTEFLAHQNCSTDPV	60
DB	1	MAHESRSSRLGVPAAGPAETGSDASEDDHPOVCAKCCAQFDPTEFLAHQNCSTDPV	60
OY	61	MYLIGQENPNNSASSEPREGHNPNQVMDTEHSPNPPSGSSVPTDPTWGPERRGEES	120
DB	61	MYLIGQENPNNSASSEPREGHNPNQVMDTEHSPNPPSGSSVPTDPTWGPERRGEES	120
OY	121	GHRLVATGTGAAGGGGLILASPKLGATPLPEESTPAPPPPPPPPPGVGSHLNIPLT	180
DB	121	GHRLVATGTGAAGGGGLILASPKLGATPLPEESTPAPPPPPPPPPGVGSHLNIPLT	180
OY	181	LEELRLVLOQROIHQOMQTEQICROVLLILSGITGAGAPSPDELGTGASSKPLPLP	240
DB	181	LEELRLVLOQROIHQOMQTEQICROVLLILSGITGAGAPSPDELGTGASSKPLPLP	240
OY	241	SPIKPVQTSKTLASSSSSSSSSGAETPKQAFPHLYHPLGSHPPSAGVGSRSHKTPAP	300
DB	241	SPIKPVQTSKTLASSSSSSSSSGAETPKQAFPHLYHPLGSHPPSAGVGSRSHKTPAP	300
OY	301	SPALPESTQOLILASPHLAFSTTGLLAAQCIGAARGLEATASPGLKPRNGSGELSTGEV	360
DB	301	SPALPESTQOLILASPHLAFSTTGLLAAQCIGAARGLEATASPGLKPRNGSGELSTGEV	360
OY	361	MGPLEKPGGRHKRCFCAKVGSALQIHLRSHTGERPYKCNVGNRPTTRGNLKVHFR	420
DB	361	MGPLEKPGGRHKRCFCAKVGSALQIHLRSHTGERPYKCNVGNRPTTRGNLKVHFR	420
OY	421	HREKTPHQMNPDPVPEHLDTVITSSGLPYGMSVPERKAEAEATPGGVERKPLVASTT	480

```
Db 421 HREKYPHQMNPHPVEHLDYITSSGLPRGMSVPPKAEAEATPGGVERKPLVASTT 480
OY 481 ALSATESLTLSTAGTATAPGLPAFNKFLMKAVEPKNADENTPGSEGSALSGVAES 540
Db 481 ALSATESLTLSTAGTATAPGLPAFNKFLMKAVEPKNADENTPGSEGSALSGVAES 540
OY 541 STATLMQLSKLTSLPSNALLTNHFKSTGSPFLPLCARALGASPSSTKLOQLVEKIDRQ 600
Db 541 STATLMQLSKLTSLPSNALLTNHFKSTGSPFLPLCARALGASPSSTKLOQLVEKIDRQ 600
OY 601 GAVAVTSAASGAPTTSAFAPSSASSGPNOCVLCRLVLSCPRALRLHYGONGGERPFCKK 660
Db 601 GAVAVTSAASGAPTTSAFAPSSASSGPNOCVLCRLVLSCPRALRLHYGONGGERPFCKK 660
OY 661 VCGRAFTSRGNLRAHFVGHKASPAARAQNSCPICQKKFTNAVTLOQHYRMHLGGQIPNGG 720
Db 661 VCGRAFTSRGNLRAHFVGHKASPAARAQNSCPICQKKFTNAVTLOQHYRMHLGGQIPNGG 720
OY 721 TALPBGGAQOENGSGFQSTVSGAGSPFOQSQSPSEELSEEEDEDEEDVDDEDS 780
Db 721 TALPBGGAQOENGSGFQSTVSGAGSPFOQSQSPSEELSEEEDEDEEDVDDEDS 780
OY 781 LAGRSESGGEKAISVRGDSEEAASGAEEVGTVAATAATAGKEMDSNEKTTQOSSLPPPPP 840
Db 781 LAGRSESGGEKAISVRGDSEEAASGAEEVGTVAATAATAGKEMDSNEKTTQOSSLPPPPP 840
OY 841 PDSLDPQPMPEGSSCVLGGKEGKPERSSPASALTPEGEATVTLVEELSLQEAAMRK 900
Db 841 PDSLDPQPMPEGSSCVLGGKEGKPERSSPASALTPEGEATVTLVEELSLQEAAMRK 900
OY 901 EPGESSSRKACEVCGAFPSQALAEHQKTHPKEGPLTVCFCRGGFLERATLKKHMLLA 960
Db 901 EPGESSSRKACEVCGAFPSQALAEHQKTHPKEGPLTVCFCRGGFLERATLKKHMLLA 960
OY 961 HHQVOPFAPHGPNIALSLVPCGSPSTSTGLSPFPRKDDPTIP 1005
Db 961 HHQVOPFAPHGPNIALSLVPCGSPSTSTGLSPFPRKDDPTIP 1005

RESULT 2
AAE17967 standard; Protein; 1005 AA.
ID AAE17967
AC AAE17967;
XX 07-MAY-2002 (first entry)
DT 07-MAY-2002 (first entry)
XX Human Sal2 protein mutant (S73C).
DE Human: proliferative disorder; tumor host range mutant virus; cancer;
KW T-HR mutant; Sal2 protein; ovarian tumour; mutant; mutcin.
XX Homo sapiens.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 73 /note= "Wild-type Ser substituted with Cys"
FT WT
XX WO200204596-A2.
XX 17-JAN-2002.
XX 05-JUL-2001: 2001MO-US21354.
XX 07-JUL-2000: 2000US-216723P.
XX 19-MAR-2001: 2001US-0812471.
XX 19-MAR-2001: 2001US-0812633.
XX (HARD ) HARVARD COLLEGE.
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
```

```
PI Benjamin TL, Li D, Mok SC, Cramer DW, Ma Y;
XX WPL: 2002-164637/21.
DR
XX Detecting protein involved in susceptibility to proliferative disease,
PT by infecting normal and abnormal proliferating cells with mutant virus,
PT detecting mutated protein allowing growth of mutant on abnormal cells
PT
XX Claim 35; Page -: 92pp; English.
XX
XX The invention relates to a method for the identification of genes
XX and their encoded proteins involved in susceptibility to proliferative
XX disorders, including cancer using a tumor host range mutant (T-HR
XX mutant) virus. The invention also provides the use of Sal2 genes and
XX mutant) virus. The invention also provides the use of Sal2 genes and
XX acquiring a proliferative disease. T-HR mutants are used to kill cancer
XX cells such as one carrying a Sal2 alteration. Transgenic and knockout
XX mouse comprising Sal2 nucleic acid are useful as research tools to
XX determine genetic and physiological features of cancer and for
XX identifying compounds that can affect ovarian and other tumours. The
XX present sequence is human Sal2 protein mutant (S73C).
XX Note: This sequence is not shown in the specification but is derived
XX from the human Sal2 wild-type protein shown as SEQ ID NO:1 (AAE1/934)
XX in page 77-79 of the specification.
XX
XX Sequence 1005 AA:
SQ
XX
XX Query Match 99.9%; Score 5272; DB 23; Length 1005;
XX Best Local Similarity 99.9%; Pred. No. 8,2e-295;
XX Matches 1004; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 MAHERSSRLGYPAGEPAELGDSBEDHPYCKCAQFTDPTFEALHONACSTDPV 60
Db 1 MAHERSSRLGYPAGEPAELGDSBEDHPYCKCAQFTDPTFEALHONACSTDPV 60
OY 61 MVLIGQOENPNNSASSEPREGHNNPQMDTEHSNPDGSSVPTDPTWGPERRGEES 120
Db 61 MVLIGQOENPNNSASSEPREGHNNPQMDTEHSNPDGSSVPTDPTWGPERRGEES 120
OY 121 GHFLVATGTAAAGGGLITLASPRLGATPLPESTPAPPPPPPPPGVSGHNLPLI 180
Db 121 GHFLVATGTAAAGGGLITLASPRLGATPLPESTPAPPPPPPPPGVSGHNLPLI 180
OY 181 LEEELRVLOQRQIHOMQMTQICRQVLLVLSLQQTGAPAPSPSELPCTGTASTRPLPLF 240
Db 181 LEEELRVLOQRQIHOMQMTQICRQVLLVLSLQQTGAPAPSPSELPCTGTASTRPLPLF 240
OY 241 SPIKPVQTSKTLASSSSSSSSSGAETPKQAFHLYHPLGSOHPFSAGVGRSHKPTPAP 300
Db 241 SPIKPVQTSKTLASSSSSSSSSGAETPKQAFHLYHPLGSOHPFSAGVGRSHKPTPAP 300
OY 301 SPALPGSTDQILASPHLAFPTTGLLAOCIGAARGLEATPAGGLKPKNGSGELSYGEV 360
Db 301 SPALPGSTDQILASPHLAFPTTGLLAOCIGAARGLEATPAGGLKPKNGSGELSYGEV 360
OY 361 MCPLEKPGGRHKCFKCAVFGSDSALOHLRSHTGEBRYKCNVCGNFTTNGNLKVHNR 420
Db 361 MCPLEKPGGRHKCFKCAVFGSDSALOHLRSHTGEBRYKCNVCGNFTTNGNLKVHNR 420
OY 421 HREKYPHQMNPHPVEHLDYITSSGLPRGMSVPPKAEAEATPGGVERKPLVASTT 480
Db 421 HREKYPHQMNPHPVEHLDYITSSGLPRGMSVPPKAEAEATPGGVERKPLVASTT 480
OY 481 ALSATESLTLSTAGTATAPGLPAFNKFLMKAVEPKNADENTPGSEGSALSGVAES 540
Db 481 ALSATESLTLSTAGTATAPGLPAFNKFLMKAVEPKNADENTPGSEGSALSGVAES 540
OY 541 STATLMQLSKLTSLPSNALLTNHFKSTGSPFLPLCARALGASPSSTKLOQLVEKIDRQ 600
Db 541 STATLMQLSKLTSLPSNALLTNHFKSTGSPFLPLCARALGASPSSTKLOQLVEKIDRQ 600
OY 601 GAVAVTSAASGAPTTSAFAPSSASSGPNOCVLCRLVLSCPRALRLHYGONGGERPFCKK 660
```

D	b		601	GAAVAVTSAAGAPTTSAPAPSSASSGPNOCYICLRVLSCPALRLHIGHGGERPFCK	660
O	y		661	VCGRAFTSGNLRAHFVGHKASPARAARONSCPICOKKFTNAVTLQOHVMHLGGQIPNGG	720
D	b		661	VCGRAFTSGNLRAHFVGHKASPARAARONSCPICOKKFTNAVTLQOHVMHLGGQIPNGG	720
O	y		721	TALPEGGGAOENSSBSTVSAGASPQQSOQSPSEELSEEDEEDEDVDIEDS	780
D	b		721	TALPEGGGAOENSSBSTVSAGASPQQSOQSPSEELSEEDEEDEDVDIEDS	780
O	y		781	LARGSESGGEKAISVGDSEBAGAEEVEVTVAATAAGKEMDENKTTOOSLPPRP	840
D	b		781	LARGSESGGEKAISVGDSEBAGAEEVEVTVAATAAGKEMDENKTTOOSLPPRP	840
O	y		841	PDSLDPQPMEOGSSGVLGKEEGEKPERSSPASALTRPEGANSVTLVEELSLQAMRK	900
D	b		841	PDSLDPQPMEOGSSGVLGKEEGEKPERSSPASALTRPEGATSVTLVEELSLQAMRK	900
O	y		901	EPGESSRKAACEVGCAFPSQAALLEHOCTHRKESPLTCVCRCGFERRATLKRMHLA	960
D	b		901	EPGESSRKAACEVGCAFPSQAALLEHOCTHRKESPLTCVCRCGFERRATLKRMHLA	960
O	y		961	HRRVQFPAPHPGNIAALSLVPGCSPSTTGLSPFRKKDPTTP	1005
D	b		961	HRRVQFPAPHPGNIAALSLVPGCSPSTTGLSPFRKKDPTTP	1005
RESULT 3					
	ID	AAL17968			
		AAL17968 standard; Protein; 1005 AA.			
X	X	AAL17968;			
X	X	07-MAY-2002 (first entry)			
X	X				
X	X	Human Sal2 protein mutant (G744P).			
X	X				
K	M	Human; proliferative disorder; tumor host range mutant virus; cancer;			
K	M	T-HR mutant; Sal2 protein; ovarian tumour; mutant; mutlein.			
O	S	Homo sapiens.			
O	S	Synthetic.			
F	T	Key	Location/Qualifiers		
F	T	Misc-difference 744	/note= "Wild-type Gly substituted with Arg"		
P	N	WO200204596-A2.			
P	D	17-JAN-2002.			
P	D				
P	E	05-JUL-2001; 2001WO-US21354.			
X	X				
P	R	07-JUL-2000; 2000US-216723P.			
P	R	19-MAR-2001; 2001US-0812471.			
P	R	19-MAR-2001; 2001US-0812633.			
X	X				
P	A	(HARD) HARVARD COLLEGE.			
P	A	(BGHM) BRIGHAM & WOMENS HOSPITAL INC.			
X	X				
P	I	Benjamin TL, Li D, Mok SC, Cramer DW, Ma Y;			
X	X				
D	R	WPI; 2002-164637/21.			
P	T				
P	T	Detecting protein involved in susceptibility to proliferative disease,			
P	T	by infecting normal and abnormal proliferating cells with mutant virus			
P	T	- detecting mutated protein allowing growth of mutant on abnormal cells			
X	X				
P	S				
C	C	Example 7; Page -: 92pp; English.			
The invention relates to a method for the identification of genes					

CC	and their encoded proteins involved in susceptibility to proliferative
CC	disorders, including cancer using a tumor host range mutant (T-HR
CC	mutant) virus. The invention also provides the use of Sal2 genes and
CC	proteins in methods of identifying a mammal having, or at a risk of
CC	acquiring a proliferative disease. T-HR mutants are used to kill cancer
CC	cells such as one carrying a Sal2 alteration. Transgenic and knockout
CC	mouse comprising Sal2 nucleic acid are useful as research tools to
CC	determine genetic and physiological features of cancer and for
CC	identifying compounds that can affect ovarian and other tumours. The
CC	present sequence is human Sal2 protein mutant (G744R).
CC	Note: This sequence is not shown in the specification but is derived
CC	from the human Sal2 wild-type protein shown as SEQ ID NO:1 (AAE17954)
CC	in page 77-79 of the specification.
XX	
SQ	Sequence 1005 AA:
Query Match	99.8%; Score 5269; DB 23; Length 1005;
Best Local Similarity	99.9%; Pred. No. 1.2e-254;
Matches 1004;	Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY	1 MAHESRSRLGYPAGEPAELGGDASEDHDPQVCAKCCAGFTDPEFLAHONACSTDPPV 60
DB	1 MAHESRSRLGYPAGEPAELGGDASEDHDPQVCAKCCAGFTDPEFLAHONACSTDPPV 60
OY	MVIIGQENNNNSASSEPRREGHNNOVMDEHSNPDDSSVPPTDTPWPERGGESS 120
DB	MVIIGQENNNNSASSEPRREGHNNOVMDEHSNPDDSSVPPTDTPWPERGGESS 120
OY	121 GHFLVAATGTAAAGGGGLILASPKLGATPLPESTPAPPPPpppppgvsgHLNIPLI 180
DB	121 GHFLVAATGTAAAGGGGLILASPKLGATPLPESTPAPPPPpppppgvsgHLNIPLI 180
OY	181 LEEIARVLQORQHOMQMTEDICROVLLLSGIVGVGPASPSELPGNTASTKPPLPLF 240
DB	181 LEEIARVLQORQHOMQMTEDICROVLLLSGIVGVGPASPSELPGNTASTKPPLPLF 240
OY	241 SPIKPVOTSKTLAASSSSSSSGAETPKOAFPHLYHPISQHPFSAGVGSRSHKPPAP 300
DB	241 SPIKPVOTSKTLAASSSSSSSGAETPKOAFPHLYHPISQHPFSAGVGSRSHKPPAP 300
OY	301 SPALPSTQOLLASPHLAFSTTGILLAAOCIGAARGLEATAPELLPKNGSSELSGEV 360
DB	301 SPALPSTQOLLASPHLAFSTTGILLAAOCIGAARGLEATAPELLPKNGSSELSGEV 360
OY	361 MGPLEKPGGRHKCRFCAKYFGSDSAIQIHLRSHTEGERPYCNVCGNFRTTGNLKVFHR 420
DB	361 MGPLEKPGGRHKCRFCAKYFGSDSAIQIHLRSHTEGERPYCNVCGNFRTTGNLKVFHR 420
OY	421 HHEKYPHYOMNHPRVEHDYVTTSGLPRGMVPRPKAEDEAATPGGVBRKLVASTT 480
DB	421 HHEKYPHYOMNHPRVEHDYVTTSGLPRGMVPRPKAEDEAATPGGVBRKLVASTT 480
OY	481 ALSATESLLLTSTAGTATAPGLPAFNKEFLVLMKAVERKKNADETPRGSGSALSVAES 540
DB	481 ALSATESLLLTSTAGTATAPGLPAFNKEFLVLMKAVERKKNADETPRGSGSALSVAES 540
OY	541 STATLMQLSKLMTSLPSMALLTNHFKSTSFPLPCARLAGASPSSETSKDOOLEKIDRQ 600
DB	541 STATLMQLSKLMTSLPSMALLTNHFKSTSFPLPCARLAGASPSSETSKDOOLEKIDRQ 600
OY	601 GAUAVTSAASGAPTTSAPRSSASAGPNOCVCLCVLSCPRALRIHYGGHGGERPKCK 660
DB	601 GAUAVTSAASGAPTTSAPRSSASAGPNOCVCLCVLSCPRALRIHYGGHGGERPKCK 660
OY	661 VCGRAFPSTRGNIRAHFVGHKASPARAONSCPICQKKTFAVTLQOHVRMLHGOIIPNGC 720
DB	661 VCGRAFPSTRGNIRAHFVGHKASPARAONSCPICQKKTFAVTLQOHVRMLHGOIIPNGC 720
OY	721 TALPFGGGAOGNGSEOSTVSGASRPQOOOSOPSPSEELSSEEEEEDEEDEEDVDDED 780
DB	721 TALPFGGGAOGNGSEOSTVSGASRPQOOOSOPSPSEELSSEEEEEDEEDEEDVDDED 780
OY	781 LAGRSSEGGEEKALISVRDSEASGAEIEGVVAAATAGKEKDSNEKTTQQGSLLPPPP 840

DB 781 LAGGSSGGKAIKSVKSDSEASAEVEGVAAATGKEKENDSNEKTTQOSSLPPPPP 840
QY 841 PDSLDPQPMEQSGSSVVGKKEEGKPERSSPASALTPGEATSVTLVELSLQEAMRK 900
DB 841 PDSLDPQPMEQSGSSVVGKKEEGKPERSSPASALTPGEATSVTLVELSLQEAMRK 900
QY 901 EPESSSRKACEVCGQAFPSQALAEHOKTHPEKPLFTCVFCROGFLERATLKKHMLLA 960
DB 901 EPESSSRKACEVCGQAFPSQALAEHOKTHPEKPLFTCVFCROGFLERATLKKHMLLA 960
QY 961 HHQVQPAHPGPNIAALSIYVGCSPSTSTGLSPPRKDDPTIP 1005
DB 961 HHQVQPAHPGPNIAALSIYVGCSPSTSTGLSPPRKDDPTIP 1005
RESULT 4
ID AAM78838 standard; Protein; 1007 AA.
AC AAM78838;
DT 06-NOV-2001 (first entry)
DE Human protein SEQ ID NO 1500.
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
XX WC200157190-A2.
PD 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX (HXSE-) HXSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI: 2001-476283/51.
DR N-PSDB: AAK51971.
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX Claim 20; Page 3790-3792; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM8020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1007 AA;
Query Match 98.0%; Score 5171; DB 22; Length 1007;
Best Local Similarity 99.5%; Pred. No. 5.3e-289;
Matches 985; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 16 GEPALGDASEEDHPQVCAKCAOFTDPTFLAHQACSTDPVMVITIGQENPNNSA 75
DB 18 GPSASENDASEEDHPQVCAKCAOFTDPTFLAHQACSTDPVMVITIGQENPNNSA 77
QY 76 SSEPRBEHNNPQVMDTEHNSNPPDSSVPTDTWGEPRGEESGHTLVATGTAAGG 135
DB 78 SSEPRBEHNNPQVMDTEHNSNPPDSSVPTDTWGEPRGEESGHTLVATGTAAGG 137
QY 136 GGLILASPKLGATPLPEESTPAPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 195
DB 138 GGLILASPKLGATPLPEESTPAPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 197
QY 196 QMTEQICRQVLLGSLGQTVGAPASPELPGTGAASSTKPLPLFSPIKVQTSKLASS 255
DB 198 QMTEQICRQVLLGSLGQTVGAPASPELPGTGAASSTKPLPLFSPIKVQTSKLASS 257
QY 256 SSSSSSSSAETPKQAFHLYHPLGSHPPSAGVGSHKRTAPSPALPGSTDOLIASP 315
DB 258 SSSSSSSSAETPKQAFHLYHPLGSHPPSAGVGSHKRTAPSPALPGSTDOLIASP 317
QY 316 HLAFFTGTGLLAOCGAAAGLEATASPGLLKPKNGSGELSYGEVMGPLEKPGRRKCRF 375
DB 318 HLAFFTGTGLLAOCGAAAGLEATASPGLLKPKNGSGELSYGEVMGPLEKPGRRKCRF 377
QY 376 CAKVEGSDALQILHLSHTEGERRYKCNVCGNRTTGNLKVHFHRRREKYPHYOMNPHV 435
DB 378 CAKVEGSDALQILHLSHTEGERRYKCNVCGNRTTGNLKVHFHRRREKYPHYOMNPHV 437
QY 436 PEHLDYVITSSGLPYGMSVPRKEAEEAATPGGVERKPLVASTALSTESTLTILSTSA 495
DB 438 PEHLDYVITSSGLPYGMSVPRKEAEEAATPGGVERKPLVASTALSTESTLTILSTSA 497
QY 496 GTATAPGLPAFNKFLVLMKAEPKAKADENTPPGSEGSASISGAESTATLMLSKLMTSL 555
DB 498 GTATAPGLPAFNKFLVLMKAEPKAKADENTPPGSEGSASISGAESTATLMLSKLMTSL 557
QY 556 PSMALLTNNHFKSTGSEFPLPCARALGASPSSETSKLQDLVEKIDROGAVAVTSAASGAPTT 615
DB 558 PSMALLTNNHFKSTGSEFPLPCARALGASPSSETSKLQDLVEKIDROGAVAVTSAASGAPTT 617
QY 616 SAPAPSSSASSGNOCYICLRVLSCPRALRLHYOGHGERPKCVCGAFTSGNMLRAH 675
DB 618 SAPAPSSSASSGNOCYICLRVLSCPRALRLHYOGHGERPKCVCGAFTSGNMLRAH 677
QY 676 FVGHKASPARAONSCPICQKFTNNAVTLQOHVNRHNLGQILPNGGTALPEGGGAOENG 735
DB 678 FVGHKASPARAONSCPICQKFTNNAVTLQOHVNRHNLGQILPNGGTALPEGGGAOENG 737
QY 736 EOSTVSGAGSFPOOQSOQSPDEBELSEEEBEDEDEEDVTDSDISLAGSGSEGEKAIS 795
DB 738 EOSTVSGAGSFPOOQSOQSPDEBELSEEEBEDEDEEDVTDSDISLAGSGSEGEKAIS 797
QY 796 VNGDSEASAEVEGVVAAATAGKENDSNEKTTQOSSLPPPPPPPSLDQPPMEQSS 855
DB 798 VNGDSEASAEVEGVVAAATAGKENDSNEKTTQOSSLPPPPPPPSLDQPPMEQSS 857
QY 856 GVLGKEEGKPERSSPASALTPGEATSVTLVELSLQEAMRKPPGSSSRKACEVCG 915
DB 858 GVLGKEEGKPERSSPASALTPGEATSVTLVELSLQEAMRKPPGSSSRKACEVCG 917
QY 916 QAFPSQALAEHOKTHPEKPLFTCVFCROGFLERATLKKHMLLAHHQVQPAHPGPN 975
DB 918 QAFPSQALAEHOKTHPEKPLFTCVFCROGFLERATLKKHMLLAHHQVQPAHPGPN 977

QY 976 AALSLVPCGSPSITSTGLSPFPKDDPTIP 1005
| | | | | | | | | | | | | | | | | | | | |
Db 978 AALSLVPCGSPSITSTGLSPFPKDDPTIP 1007

RESULT 5
AAM79822

ID AAM79822 standard; Protein; 1019 AA.
XX AAM79822;
XX
DT 06-NOV-2001 (first entry)
DE Human protein SEQ ID NO 3468.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorder; arthritis; inflammation.
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
PI Tang YH, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XZ WPI; 2001-476283/51.
DR N-PSDB; AAK52955.
DR
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 347-348; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK51435) and the
CC encoded polypeptides (AAM78323-AAAM80102) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1019 AA;

Query Match 97.8%; Score 5161; DB 22; Length 1019;
Best Local Similarity 99.3%; Pred. No. 2e-288;
Matches 983; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

[illegible]

AC AAE17955;
XX
DT 07-MAY-2002 (first entry)
XX
DE Mouse Sal2 protein.
XX
KM Mouse; proliferative disorder; tumor host range mutant virus; cancer;
KM T-HR mutant; Sal2 protein; ovarian tumour.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT Misc-difference 252
FT /note= "Encoded by CTC"
XX
PN W0200204596-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US21354.
XX
PR 07-JUL-2000; 2000US-216723P.
PR 19-MAR-2001; 2001US-0812471.
PR 19-MAR-2001; 2001US-0812633.
XX
PA (HARD) HARVARD COLLEGE.
PA (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
XX
PI Benjamin TL, LI D, Mok SC, Cramer DW, Ma Y;
XX
DR MPI: 2002-164637/21.
DR N-PSDB: AAD28652.
XX
PT Detecting protein involved in susceptibility to proliferative disease,
PT by infecting normal and abnormal proliferating cells with mutant virus,
PT detecting mutated protein allowing growth of mutant on abnormal cells
XX
PS Disclosure: Page 85-87; 92pp: English.
XX
CC The invention relates to a method for the identification of genes
CC and their encoded proteins involved in susceptibility to proliferative
CC disorders, including cancer using a tumor host range mutant (T-HR
CC mutant) virus. The invention also provides the use of Sal2 genes and
CC proteins in methods of identifying a mammal having, or at a risk of
CC acquiring a proliferative disease. T-HR mutants are used to kill cancer
CC cells such as one carrying a Sal2 alteration. Transgenic and knockout
CC mouse comprising Sal2 nucleic acid are useful as research tools to
CC determine genetic and physiological features of cancer and for
CC identifying compounds that can affect ovarian and other tumours. The
CC present sequence is mouse Sal2 protein.
XX
SQ Sequence 1002 AA:

Query Match 88.6%; Score 4677.5; DB 23; Length 1002;
Best Local Similarity 89.0%; Pred. No. 1.2e-260;
Matches 894; Conservative 25; Mismatches 83; Indels 3; Gaps 3;

QY 1 MAHESRRSLRGYPAGEPALGSDASEDHQVCAKCAQFDPTEFLAHQNACTDDPV 60
DB 1 MAQETGSSSRLLGPGCEPARGGDASEHHPOVCAKCAQFDPTEFLAHQNSCTDDPV 60
QY 61 AVIIGQENPNNSASASEPPEGHNNPOVMDTEHNSPPDSGSSVPTDPTWGPGRGESS 120
DB 61 MYIIGQENPNNSASASAPPEGHNSRQVMDTEHNSPPDSGSSGPPDPTWGPGRGESS 120
QY 121 GFPLVAATGTAAAGGGGLIASPRLGATPLPEESTPAPPPPPPPPGVSGHLNPLI 180
DB 121 GGFVLVATGTAAAGGGGLIASPRLGATPLPEESTPAPPPPPPPPGVSGHLNPLI 180
QY 181 LDELRLVLOQRQIHQMOMTEQICRQVLLLSLGQYTGARASPELPEGTGASSTKPLPLF 240
DB 181 LDELRLVLOQRQIHQMOMTEQICRQVLLLSLGQYTGARASPELPEGTGASSTKPLPLF 240

QY 241 SPIKPVQTSKTLASSSSSSSGAEPTKQAFHLVHPLGSHQHPFSAGVGSHKPTPAP 300
DB 241 SPIKAQCTKTTA-SSSSSSSSGAEPKQAFHLVHPLGSHQHPFSAGVGSHKPTPAP 299
QY 301 SPALPGSTQDLIASPHLAPPTTGLLAOCLGAARGLSATSPGLLKRKNGSGELSGYV 360
DB 300 SPALPGSTQDLIASPHLAPPTTGLLAOCLGAARGLSATSPGLLKRKNGSGELSGYV 359
QY 361 MGPLEKPGGRHKCRCAKAVFGSDALQIHLRSHNGEPRYKCNVCGNRRFTTRNLKVHFR 420
DB 360 ISSLEKPGGRHKCRCAKAVFGSDALQIHLRSHNGEPRYKCNVCGNRRFTTRNLKVHFR 419
QY 421 HREKYPHVOMNPNRVEHLDVYITSSGLPYGMSVPEPEKAEBAATPGGVERKPLVASTT 480
DB 420 HREKYPHVOMNPNRVEHLDVYITSSGLPYGMSVPEPEKAEBAATPGGVERKPLVASTT 479
QY 481 ALSATESLTLSTAGTATAPGLPAFENKFVLMKAVEPKKADENTPPGSEGAISGVAES 540
DB 480 ALSATESLTLSTAGTATAPGLPAFENKFVLMKAVEPKKADENTPPGSEGAISGVAES 539
QY 541 STATLMQSLKMLSLPSPWALLTNHFKSTGSRPLPLCARALGASPSETSLOOLVEKIDRQ 600
DB 540 GSATRMQSLKMLSLPSPWALLTNHFKSTGSRPFVLEPLGASPSETSLOOLVEKIDRQ 599
QY 601 GAVAVTSAASGAPPTAPAPSSSASGPNOCVICLRVLSCPRALRLHYGOHGERPFKCK 660
DB 600 GAVAVTSAASGAPPTAPAPSSSASGPNOCVICLRVLSCPRALRLHYGOHGERPFKCK 658
QY 661 VCGRAFTSGNLRHAHFVGHKASPARAQNCSPICOKFTNAVTLOQHVBMHLGGQIPNGG 720
DB 659 VCGRAFTSGNLRHAHFVGHKASPARAQNCSPICOKFTNAVTLOQHVBMHLGGQIPNGG 718
QY 721 TALPEGGGAQENGSRQSTYSGAGSFPOQSOQSPSEBELSEEEDEDEEDVDYDEDS 780
DB 719 SALSIEGGGAQENGSRQSTYSGAGSFPOQSOQSPSEBELSEEEDEDEEDVDYDEDS 777
QY 781 LAGRGSESGEKAISVNGDSEASGAEVEGTVAAATAGKEMDSNEKTTQOSSLPPPP 840
DB 778 LAGRGSESGEKAISVNGDSEASGAEVEGTVAAATAGKEMDSNEKTTQOSSLPPPP 837
QY 841 PSLDPOPEMEGSSGVLGKEGKPEKRRSSPASALRPEGATSVTLVEELSLOEAMRK 900
DB 838 PNLIDHPQPEMEGSSGVLGKEGKPEKRRSSPASALRPEGATSVTLVEELSLOEAMRK 897
QY 901 EPGESSRKAACEVCGAOFPSOALEEHOKTHPKREGPLFTVCFCROGFLERATLKKHMLA 960
DB 898 DPESSGRKACEVCGAOFPSOALEEHOKTHPKREGPLFTVCFCROGFLERATLKKHMLA 957
QY 961 HHQVFPFAPHPONIALSLVPGCSPSITSTGLSPFPRKDDPTIP 1005
DB 958 HHQVFPFAPHPONIALSLVPGCSPSITSTGLSPFPRKDDPTIP 1002

RESULT 7
AAB93193
ID AAB93193 standard; Protein: 813 AA.
XX
AC AAB93193;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12150.
XX
KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.


```
KM epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN MO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR:
XX
DR WPI: 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
PS Example 4; SEQ ID NO: 28963; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention.
XX
SQ Sequence 330 AA;
XX
Query Match 9.3%; Score 489; DB 22; Length 330;
Best Local Similarity 33.3%; Pred. No. 2,1e-20;
Matches 186; Conservative 51; Mismatches 114; Indels 108; Gaps 14;
XX
OY 529 SEGSAISGVAESSTATIMQLSKMLTSLPSWALLTNHFKSTGSPFLPLCARALGASPSSETS 588
DB 3 SSPAADCGPAGSATTTFTNPL-----LP---LMSEQFKA---KPPFGGLDLSAQASSETS 49
XX
OY 589 KLOQLVEKIDROGAVANTSAASGAPTTAPAPSSASSGPNQVCYLRLVSCPRALRLHY 648
DB 50 KLOQLVENIDRK-----ATDPNCCILCHRYLSCQSALAKMHY 85
XX
OY 649 GGGGERPFPKVCVCGRAFTSGNLRAHFVGHKASPAARAQNSCPICQKFTNAVTLQOHV 708
DB 86 RHTGTGERPFPKCKICGRAFTTKGNLKTHTSVHRAMPPLRVQHCSPICQKFTNAVVVQOHI 145
XX
OY 709 RMLHGOIPINGGTALPEGGGAQOENGSEQSVSGAGSPPOQSOOPSPBEELSEEEBED 768
DB 146 RMLHMGQIPIN--TPVPDYSSESM-----SDTGSF-----DEKNFDD 180
XX
OY 769 EEEEDVTDDEDSLARGSESGEKAISVRGDSSEASGAEEVEGTVAATAATACKEMPSNEK 828
DB 181 LDNFSDENMED--CPESGIPTDPKASADASQDLSSSPLPLEMSSTIALENNQKMINAAGLA 238
XX
OY 829 TTQSSSLPPPPDSDLQOPQPMEOGS-----SGVLG---KEEGKPERSSSPAS- 875
DB 839 EOLQASL-----KSVENGSIEGDVLITNDSSSVGGMESQASQSPAISESTSSM 286
XX
OY 876 -ALTEGGEATSVTLVEELSLQELAMRKPEGESSSRKACCEVCGAPPSQAA 923
DB 287 QALSPSNS-----TOEFHKSPTIEEKP-----QRAVPSSEA 317
XX
RESULT 15
```

```
AAM69245
ID AAM69245 standard; Protein; 330 AA.
XX
AC AAM69245;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 29551.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN MO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR:
XX
DR WPI: 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 29551; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
SQ Sequence 330 AA;
XX
Query Match 9.3%; Score 489; DB 22; Length 330;
Best Local Similarity 33.3%; Pred. No. 2,1e-20;
Matches 136; Conservative 51; Mismatches 114; Indels 108; Gaps 14;
XX
OY 529 SEGSAISGVAESSTATIMQLSKMLTSLPSWALLTNHFKSTGSPFLPLCARALGASPSSETS 588
DB 3 SSPAADCGPAGSATTTFTNPL-----LP---LMSEQFKA---KPPFGGLDLSAQASSETS 49
XX
OY 589 KLOQLVEKIDROGAVANTSAASGAPTTAPAPSSASSGPNQVCYLRLVSCPRALRLHY 648
DB 50 KLOQLVENIDRK-----ATDPNCCILCHRYLSCQSALAKMHY 85
XX
OY 649 GGGGERPFPKVCVCGRAFTSGNLRAHFVGHKASPAARAQNSCPICQKFTNAVTLQOHV 708
DB 86 RHTGTGERPFPKCKICGRAFTTKGNLKTHTSVHRAMPPLRVQHCSPICQKFTNAVVVQOHI 145
XX
OY 709 RMLHGOIPINGGTALPEGGGAQOENGSEQSVSGAGSPPOQSOOPSPBEELSEEEBED 768
DB 146 RMLHMGQIPIN--TPVPDYSSESM-----SDTGSF-----DEKNFDD 180
XX
OY 769 EEEEDVTDDEDSLARGSESGEKAISVRGDSSEASGAEEVEGTVAATAATACKEMPSNEK 828
DB 181 LDNFSDENMED--CPESGIPTDPKASADASQDLSSSPLPLEMSSTIALENNQKMINAAGLA 238
XX
OY 829 TTQSSSLPPPPDSDLQOPQPMEOGS-----SGVLG---KEEGKPERSSSPAS- 875
```



```
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-483446/52.
XX DR
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX PS
XX PS Example 4; SEQ ID NO: 30291; 650pp + Sequence listing; English.
CC CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX SQ
XX Sequence 336 AA; |
XX
Query Match 8.4%; Score 441; DB 22; Length 336;
Best Local Similarity 34.3%; Pred. No. 1.2e-17;
Matches 119; Conservative 41; Mismatches 89; Indels 98; Gaps 11;
QY 585 SETSKLOOLEKIDROGAVAVTSAASGAPTTAPAPSSSSAGPNOCVICRLVLCPRAL 644
DB 62 SETLKLQOLEVENDK-----ATDPNCECLICHRVLSGSSSL 97
QY 645 RLHYGOGHGERPRCKVCYKCGRAFSTRGNLRHFVGHKASPAARAQNSCPICQKKFTNAVTL 704
DB 98 KMHYRTHYGERPRFOCKICGRAFSKGNLKTGLGVHRTNTSIKTHSCPICQKKFTNAVML 157
QY 705 QOHVRMHLGQIPIPGCTALPPEGGAQAQENGSPQSTVSGAGSPPOQSOQSPPEELSEEE 764
DB 158 QOHIRNMHGGQIIPN--TPLE--NPCDFTGSEPMTVGENGS-----TGAICHDH 202
QY 765 EEEDEEEEDVTDSDSLAGSGESGGEKALISVRGDSEASGAEEVGTV-----AAATA 819
DB 203 VIESIVVE-----VSSQEARPSSSSKVPPLPSTHSASPTL 238
QY 820 GKEMDSNEKTKTOQSSLP-----PPPSLDOPQPEOG-----SSGVLGKEEGG 865
DB 239 GFAM-----MASIDAPGKVGPAFNLQROGSRENGSVESDGLTNDSSSLMGDQ--- 287
QY 866 KPERSSSPASALTPEGEATSVTLIVEELSLQEMARKPEPGSSSRKACE 912
DB 288 --YQSRSPDILETTTFQALSPANSQAESIKS---KSPDAGSRAESSE 329
XX
RESULT 23
XX ID AAM70641 standard; Protein; 336 AA.
XX AC AAM70641;
XX DT
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30947.
XX KW Human: bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
```

```
XX XX
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00668.
XX PF
XX XX 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488900/53.
XX DR
XX XX Human genome-derived single exon nucleic acid probes useful for
XX XX analyzing gene expression in human bone marrow -
XX PT
XX PS Example 4; SEQ ID NO: 30947; 658pp + Sequence listing; English.
XX PS
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention.
XX SQ
XX Sequence 336 AA; |
XX
Query Match 8.4%; Score 441; DB 22; Length 336;
Best Local Similarity 34.3%; Pred. No. 1.2e-17;
Matches 119; Conservative 41; Mismatches 89; Indels 98; Gaps 11;
QY 585 SETSKLOOLEKIDROGAVAVTSAASGAPTTAPAPSSSSAGPNOCVICRLVLCPRAL 644
DB 62 SETLKLQOLEVENDK-----ATDPNCECLICHRVLSGSSSL 97
QY 645 RLHYGOGHGERPRCKVCYKCGRAFSTRGNLRHFVGHKASPAARAQNSCPICQKKFTNAVTL 704
DB 98 KMHYRTHYGERPRFOCKICGRAFSKGNLKTGLGVHRTNTSIKTHSCPICQKKFTNAVML 157
QY 705 QOHVRMHLGQIPIPGCTALPPEGGAQAQENGSPQSTVSGAGSPPOQSOQSPPEELSEEE 764
DB 158 QOHIRNMHGGQIIPN--TPLE--NPCDFTGSEPMTVGENGS-----TGAICHDH 202
QY 765 EEEDEEEEDVTDSDSLAGSGESGGEKALISVRGDSEASGAEEVGTV-----AAATA 819
DB 203 VIESIVVE-----VSSQEARPSSSSKVPPLPSTHSASPTL 238
QY 820 GKEMDSNEKTKTOQSSLP-----PPPSLDOPQPEOG-----SSGVLGKEEGG 865
DB 239 GFAM-----MASIDAPGKVGPAFNLQROGSRENGSVESDGLTNDSSSLMGDQ--- 287
QY 866 KPERSSSPASALTPEGEATSVTLIVEELSLQEMARKPEPGSSSRKACE 912
DB 288 --YQSRSPDILETTTFQALSPANSQAESIKS---KSPDAGSRAESSE 329
XX
RESULT 24
XX ID AAM18485 standard; Protein; 336 AA.
XX AC AAM18485;
XX DT
XX DT 12-OCT-2001 (first entry)
XX DE Peptide #4919 encoded by probe for measuring cervical gene expression.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
```

KM cervical cancer.
XX Homo sapiens.
OS
XX WO200157278-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00670.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR:
PI
XX WPI; 2001-488901/53.
XX
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
PS
XX
XX Claim 27; SEQ ID No 23311; 487bp; English.
CC
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs; see A110068-A18459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 336 AA:
SQ
Query Match 8.4%; Score 441; DB 22; Length 336;
Best Local Similarity 34.3%; Pred. No. 1.2e-17;
Matches 119; Conservative 41; Mismatches 89; Indels 98; Gaps 11;
OY 585 SETSKLQQLVEKIDROGAVAVTSAASGAPPTTSAPAPSSASSGPNQVCYLRLVSCPRAL 644
DB 62 SETKLQQLVEKIDK-----ATTDPECLICHRYLSCQSSL 97
OY 645 RLHYGOHGERPPKCKVCGRAFTSGNLRAHFVGHKASPARAONSCPTCQKKFTNAVTL 704
DB 98 KMHYRTHGERTPOCKICGRAFTSGNLKTHLGVHRTNTSITQHSCPTCQKKFTNAVML 157
OY 705 QOHVAMHLGQIPNGGTALPEGGAAGQENGSEGSTVSGASFPQOQSOPSPPEELSEEE 764
DB 158 QOHIRMHMGQIPN--TPLE--NPCDFTGSEPMIVGENG-----TGAICHDD 202
OY 765 EEEDEEEDVDYDSDSLAGRSGESGGEKAISVRGDSEASGAEEVGTV-----AAATA 819
DB 203 VIESIDVEE-----VSSQEAPESSSSKVPTPLPSIHSASPTL 238
OY 820 GKEMDSNEKTTQOSSLPPP-----PPPSLDOPQPMEOG-----SSGVLGKEEGG 865
DB 239 GFAM-----MASLDAPGVGPAPFNLDQROGSRENGVSVESDGLTNDSSSLMGDOE--- 287
OY 866 KPERSSPASALTPPEGATSVTLVEELSLQEAMRKREPESSSRKACE 912
DB 288 ---YQSHSPDILETTTSFOALSPANSQAESIKS---KSPDAGKAESSE 329
RESULT 25

AAM30956
ID AAM30956 standard; Protein; 336 AA.
XX
XX AAM30956;
AC
XX
XX 17-OCT-2001 (first entry)
DT
XX
XX Peptide #4993 encoded by probe for measuring placental gene expression.
DE
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200157272-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00663.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR:
PI
XX WPI; 2001-488897/53.
XX
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
PS
XX
XX Claim 27; SEQ ID No 31225; 654bp; English.
CC
XX The present invention relates to single exon nucleic acid probes (SENPs;
CC see A131315-A157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
XX Sequence 336 AA:
SQ
Query Match 8.4%; Score 441; DB 22; Length 336;
Best Local Similarity 34.3%; Pred. No. 1.2e-17;
Matches 119; Conservative 41; Mismatches 89; Indels 98; Gaps 11;
OY 585 SETSKLQQLVEKIDROGAVAVTSAASGAPPTTSAPAPSSASSGPNQVCYLRLVSCPRAL 644
DB 62 SETKLQQLVEKIDK-----ATTDPECLICHRYLSCQSSL 97
OY 645 RLHYGOHGERPPKCKVCGRAFTSGNLRAHFVGHKASPARAONSCPTCQKKFTNAVTL 704
DB 98 KMHYRTHGERTPOCKICGRAFTSGNLKTHLGVHRTNTSITQHSCPTCQKKFTNAVML 157
OY 705 QOHVAMHLGQIPNGGTALPEGGAAGQENGSEGSTVSGASFPQOQSOPSPPEELSEEE 764
DB 158 QOHIRMHMGQIPN--TPLE--NPCDFTGSEPMIVGENG-----TGAICHDD 202
OY 765 EEEDEEEDVDYDSDSLAGRSGESGGEKAISVRGDSEASGAEEVGTV-----AAATA 819
DB 203 VIESIDVEE-----VSSQEAPESSSSKVPTPLPSIHSASPTL 238
OY 820 GKEMDSNEKTTQOSSLPPP-----PPPSLDOPQPMEOG-----SSGVLGKEEGG 865
DB 239 GFAM-----MASLDAPGVGPAPFNLDQROGSRENGVSVESDGLTNDSSSLMGDOE--- 287
OY 866 KPERSSPASALTPPEGATSVTLVEELSLQEAMRKREPESSSRKACE 912

Db 288 --YQSRSPDILETTSFQALSPANQAESIKS---KSPDAGSKAESSE 329

RESULT 26

AA06075

ID AA06075 standard; Protein; 336 AA.

XX

AC AA06075;

XX

DT 09-OCT-2001 (first entry)

DE Peptide #4757 encoded by probe for measuring breast gene expression.

XX

KW Probe; human; breast disease; breast cancer; development disorder; inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX

OS Homo sapiens.

XX

PN WO200157270-A2.

XX

PD 09-AUG-2001.

XX

PF 29-JAN-2001; 2001WO-US00661.

XX

PR 04-FEB-2000; 2000US-0180312.

XX

PR 26-MAY-2000; 2000US-0207456.

XX

PR 30-JUN-2000; 2000US-0608408.

XX

PR 03-AUG-2000; 2000US-0632366.

XX

PR 21-SEP-2000; 2000US-0234687.

XX

PR 27-SEP-2000; 2000US-0236359.

XX

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI: 2001-476286/51.

XX

PT Novel single exon nucleic acid probe used to measuring gene expression

XX

PT in a human breast -

PS Claim 27; SEQ ID No 14815; 322pp; English.

XX

CC The present invention relates to novel single exon nucleic acid probes

CC (see AA100010-AA110067). The present sequence is a peptide encoded by one

CC such probe. The probes are useful for measuring human gene expression in

CC a human breast sample, where the probe hybridises at high stringency to a

CC nucleic acid expressed in the human breast. The probes are useful for

CC predicting, diagnosing, grading, staging, monitoring and prognosing

CC diseases of the human breast, particularly those diseases with polygenic

CC aetiology. The diseases include: breast cancer; disorders of development,

CC inflammatory diseases of the breast; fibrocystic changes, proliferative

CC breast disease and non-carcinoma tumours.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

XX

SO Sequence 336 AA;

Query Match 8.4%; Score 441; DB 22; Length 336;

Best Local Similarity 34.3%; Pred. No. 1.2e-17;

Matches 119; Conservative 41; Mismatches 89; Indels 98; Gaps 11;

QY 585 SETSKLOQLEVERKIDRGAVAVTSAASGAPRTSAPAPSSASGPMQVCVLRVLSGPRAL 644

Db 62 SETKLQQLQVENIDK-----ATTDPMECLTCHRVLSQSSSL 97

QY 645 RLHYGQHGGERPFCCKVCGRAFSTRGNLRAHFVGHKASPAARAONSCPTCKKFTNAVTL 704

Db 98 KMHYRHTGERPQCKICGAFSTKGNLKHVLGVRHTNTSITQNSCPTCKKFTNAVNL 157

QY 705 QQHVMHLLGGQIPNGSTALPEGGGAOENSGESTVSGAGSFPQOOSQPPSEELSEEE 764

Db 158 QQHIRMHGGQIPN--TLPE--NPCDFTGSEPMTCGENS-----TCAICDD 202

QY 765 EEDDEEEDVTDSDSLAGRGSESGEKAI SVRGDSEEAAGAEVGTV-----AAATA 819

Db 203 VIESIDVEE-----VSSQEAAPSSSKVPTPLPSIHASPTL 238

QY 820 GKENDSNEKTTQOSSLPP-----PPDSLDQOPMEOG-----SSCVLGKKEGG 865

Db 239 GFAM-----MASLDAPGKYGPAPFNLQROGSRENGVSVDGLTNDSSSLMGDOE--- 287

QY 866 KPERSSPASALTFEGFATSVTLVEELSLQEAARKKEGESSRRACE 912

Db 288 --YQSRSPDILETTSFQALSPANQAESIKS---KSPDAGSKAESSE 329

RESULT 27

ABG40328

ID ABG40328 standard; Peptide; 336 AA.

XX

AC ABG40328;

XX

DT 19-AUG-2002 (first entry)

XX

DE Human peptide encoded by genome-derived single exon probe SEQ ID 29993.

XX

KW Human; single exon probe; asthma; lung cancer; COPD; IHD;

XX

KW Chronic obstructive pulmonary disease; interstitial lung disease;

XX

KW familial idiopathic pulmonary fibrosis; neurofibromatosis;

XX

KW tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;

XX

KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

XX

KW pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome;

XX

KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

XX

KW primary ciliary dyskinesia; pulmonary hypertension;

XX

KW hyaline membrane disease.

XX

OS Homo sapiens.

XX

PN WO200186003-A2.

XX

PD 15-NOV-2001.

XX

PF 30-JAN-2001; 2001WO-US00665.

XX

PR 04-FEB-2000; 2000US-180312P.

XX

PR 26-MAY-2000; 2000US-207456P.

XX

PR 30-JUN-2000; 2000US-0608408.

XX

PR 03-AUG-2000; 2000US-0632366.

XX

PR 21-SEP-2000; 2000US-0234687P.

XX

PR 27-SEP-2000; 2000US-0236359P.

XX

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI: 2002-114183/15.

XX

PT Spatially-addressable set of single exon nucleic acid probes, used to

XX

PT measure gene expression in human lung samples -

XX

PS Claim 27; SEQ ID No 29993; 634pp; English.

XX

CC The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived

CC from human lung comprising single exon nucleic acid probes having one of

CC 12614 nucleic acid sequences mentioned in the specification, or their

CC complements or the 12387 open reading frames derived from the 12614

CC probes. Also included are a microarray comprising the novel set of

CC probes; the novel set of probes which hybridise at high stringency to a

CC nucleic acid expressed in the human lung; measuring gene expression in a

CC sample derived from human lung, comprising (a) contacting the array with

CC a collection of detectably labeled nucleic acids derived from human lung

Db	481	CECCCKGFTSWRSNLHMQHVRHSGEKPK-----CECCDKSFSQAIDFRVHQRVHTGKPK	536
Oy	719	GGTALPEGGGAAGGSEOSTVSGAGSPFQQOQSOQPSPEELSEEDEEEDVDE	778
Db	537	CGVC-----KGFGEOSGLSHQRVHTGKPK-----K	564
Oy	779	DSLGRGSESGSEKALSVNRDSEASGAEEVGVVAAAATACKEMDSNEKTTQOSSLPP	838
Db	565	CDVCKGKGRYSQFIYHQKHTEKPKYKCECG-----KGFERST-----	604
Oy	839	PPPSLDLPQPMEOGSSGVLGKKEGKPERSSSPASALTPEGATSVTLVEELTQEAM	898
Db	605	-----NLRRHQFVHTGKPKPHL-CECGK-----AFSLPSNL	634
Oy	899	RKEFGESSRK--ACGVCGQAFPSQALAEHQKTHPKGPRFTCYFCRGGFLERAILKKH	956
Db	635	RVLHGVHTREKLEKCECGKGFSSQARLEAHQVRHTEGKP-YKCDICDKDFRRHSRLTYH	693
RESULT	29		
AAB21003			
ID	AAB21003	standard; Protein; 799 AA.	
AC	AAB21003;		
XX			
DT	19-DEC-2000	(first entry)	
XX			
DE	Human nucleic acid-binding protein, NUABP-7.		
XX			
KM	Human nucleic acid-binding protein; NUABP; agonist; antagonist; EST;		
KM	expressed sequence tag; drug screening; recombinant expression; antibody		
KW	reproductive disorder; infertility; immunological disorder;		
XX	neurological disorder; cell proliferative disorder; cancer; tumour.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200044900-A2.		
XX			
PD	03-AUG-2000.		
XX			
XX	28-JAN-2000; 2000WO-US02237.		
PF			
PR	29-JAN-1999; 99US-0117904.		
PR	29-JAN-1999; 99US-0117905.		
XX			
PA	(INCY-) INCYTE PHARM INC.		
PI	Tang YT, Lal P, Hillman JL, Yue H, Azimzal Y, Lu AMD, Baughn MR;		
PI	Tran B, Shih IL, Au-Young JL,		
XX			
XX	WPI: 2000-499332/44.		
DR	N-PSDB; AAA72386.		
XX			
PT	Novel nucleic acid binding proteins, used to identify agonists and		
PT	antagonists of them, for the treatment of reproductive, immunological,		
PT	neurological and cell proliferative disorders including cancer -		
XX			
PS	Claim 1; Page 98-100; 180pp; English.		
XX			
CC	Sequences AAB20997-B21051 represent novel human nucleic acid-binding		
CC	proteins (NUABPs) which are encoded by the cDNA sequences		
CC	AA72382-A72436. The cDNAs were produced by extension from an appropriate		
CC	EST (expressed sequence tag) using primers designed using the EST. The		
CC	invention also relates to expression constructs, host cells and		
CC	transgenic organisms comprising a human NUABP nucleic acid, recombinant		
CC	production of the human NUABPs, and antibodies against the human NUABPs,		
CC	and also to methods of screening modulators of human NUABP activity or		
CC	expression. The human NUABPs, and their agonists and antagonists are used		
CC	to treat diseases associated with overexpression or underexpression of		
CC	functional NUABPs. Human NUABP proteins and nucleotides, and NUABP		
CC	agonists and antagonists can be used to diagnose, treat and prevent		
CC	reproductive, immunological, neurological and cell proliferative		
CC	disorders. Reproductive disorders that may be treated using compositions		

Query Match	Best Local Similarity	Matches 160;	Conservative 66;	Mismatches 280;	Indels 214;	Gaps 23;
SO Sequence	799 AA;					
QY 278	PLGSO--HPFSAGVGRSHKPTPARPALPGSDIQLASPHLAPSTTGLLAOOLGARG	336				
DB 244	PLCEKPRPGCEGCGKGFST-----SPRLPLRN--VHTGCKCSOSSHLTTHQIHPEK	295				
QY 337	LEATASDGLLKPR-----NSGELSYGEVWGPLEKPGGRHKRCRCARVFGSDSALQI	388				
DB 296	LMRCHESGDCEFNKSSFSHSYOSNHTGEKSY-----KDCGCGKGFSSSTGLII	341				
QY 389	HLRSHGGERYKCKVNCGRFTTGNLKYHFNHREKPYHYOMPHRYPHRIDYITTSGL	448				
DB 342	HTHTHGEKPYKCECGKCFSSQSNFOCHQVRHTEKRP-----YKCECGK	387				
QY 449	PYGMVY-----PRKAEDEAATPGGVEKRPVASTALATSELTTLSTAG	496				
DB 388	GFGMSVNLRVHQRHGEKPYKCE-----	412				
QY 497	TATAPGLPANKRYLMAV--EPKNKADENTPQSGESAISVABSSTVTLMOLSLMT	553				
DB 413	--CGKFTQAHFHQVRHVTGEKPYKCD-----VCGKGFSSNPLCHRRVHT	459				
QY 554	SLPSMAL-----LTNHFK-STGSPFLPICARALGASPSFSKLOQLVEKIDRQ	600				
DB 460	GKPYKCECGCKGFTRTMDLHIFRVHTGEKPYK--CKCGKGFSSQSNLQVHONVTGE	517				
QY 601	GAVAVYSAASGADPTSPADPSSASSG--PNOCVTLRLVLSCPRALRLHYGOGGERPFK	658				
DB 518	KRFKCECTCGKGFSSQSLQTHQVRHVTGEKPYKCDVCGKDFSSNKLHGVHTGEKPYK	577				
QY 659	CRVCGEAFSTRGNLRAHFVGHKASPAANAQNSCPICQCKFTNAVYTLQOHVHMLGGQIPN	718				
DB 578	CECGGKGFSSKRNLMHQRVHSGEKPYK--CEQCDSSFSQALDRVHQRVHTGEKPYK	633				
QY 719	GGTALPEGGGAOENSGEOSTVSGAGSEFPQOOSQSPSEELSEEEDEEEDVETDE	778				
DB 634	CGVY--GKGFSSQSGILOSHQVRHVTGEKPY-----K	661				
QY 779	DLAAGSGSESGGKALSVYRQDSEASGAEFEVGYTAAATAGKENDSNEKTQOOSLPP	838				
DB 662	CDVCGGGERYSQFIYHQRHHTGEKPYKCECG--KGFERSL-----	701				
QY 839	PPDSDIDQOPQMGSSGVLGKKEGCKPERSSSPASALTPEGEATSVTLVEELSLQEAM	898				
DB 702	----NLRHHQRVHTGEKPHI--CECGK-----AFSLPSNL	731				
QY 899	RKEPGESSSRK--ACEVCGQAFPSOALAEHQTHPKRGRLTCVFCROGFLERATLKKH	956				
DB 732	RVHLGVHTREKLEFKCECGKGFSSQARLEAHQRVHTGEKPYKCDICDKDFRHSRLTYH	790				
RESULT 30						
AAAB0283						
ID	AAAB0283 standard; Protein; 927 AA.					
XX	AAAB0283;					

[illegible]

XX primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
PS Claim 8; SEQ ID 15609; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 498 AA:
Query Match 6.2%; Score 329.5; DB 22; Length 498;
Best Local Similarity 22.0%; Pred. No. 5e-11;
Matches 146; Conservative 70; Mismatches 240; Indels 209; Gaps 24;
OY 352 SGETSYGEVWGRPLEKPGGRHKRCFAKVGSGDSALDIHLRSHGERYKVCNRFETR 411
DB 14 AGELPF-----RCANCPKAGALSKLKHONGHGERYACADCKSKSFDP 59
OY 412 GNLKHFHRHREKYHVOMNPRVPEHLDYVITSSGLPYGMSVPERKAEERAPGGGVE 471
DB 60 S-----VERKHNR--THAQLR-----YSCERCGKAYV-ELKDLANHESSHT-----G 99
OY 472 RKPLVAST--TALSATESLTLLSTAGTATAPGLPAPNK-FVLKKAVERPKKADENTPPG 528
DB 100 ERPEFLCECGKSFSSSLTCHQRIHAQKPYRCAPCGKGTOLSSYQSHERTISGEKP- 158
OY 529 SEGSAIGVAVESSTATLMQLSKMTSLPSWALLTNHFKTSGSPFLPCARLGLGASPSETS 588
DB 159 -----FLLCPRCGR-MFSDPSFR 175
OY 589 KLOQLVEKID-----ROGAVAVATSAAGAPTTSAPADSSASSGPNOCTICLRV 637
DB 176 RHQAHGKGVKRYHCKGCKDFRQ-----PADLAMHRRVHNGDRFKICQCKT 223
OY 638 LSCPRALRLHYGNGGERPFKCKVCGRAFSTRGNLRAHYGNKASPARAQNSCPICOKK 697
DB 224 FVASMIDKRLHNLVHSGORPFCRCECGRAFERASLTGHSRVH---SGERPFHCNACGS 279
OY 698 FTNVTATLOOHVBMHNGGQ---IPNG-----GTALPEGGAAOENGSEOSTVSGAGSPQQ 749
DB 280 FVASSSLKKHERTHSSSEAGVPPROELVVGALPV--GVAEGSSAPARAAGALDDPPR- 336
OY 750 QSOQPSPEELSEEEEDDEEEDVTDSDSLAGRSGSESGEKA-----ISVNGDSEB 802
DB 337 -----GLGLPESSGVMTQWGVGVTVHVEQ 366
OY 803 ASGAEVEYGTAAATAATAGKEMDSNEKTTQOSSLRPPPPPSLDOPQPMEGSSGVLGCKE 862
DB 367 DAGVREAPGLEGAGEAGE-BADEK-----PPOFVCR-ECKETFTSTMTLLRRH 413

OY 863 EGGKPERSSPASALTPGEATSVTLVEELSLQEAARKPEGESSRRKACEGGAFFPSQA 922
DB 414 ERSHBELRPP-----CTQCGKSFSDRA 436
OY 923 ALDEHOKTHPEKGFLETCVFCRQGLERATLKKHMLAHNQVOPFAPHGQNALSLP 982
DB 437 GLRKHSRTHSSVPR-YTCRPHCKAFSLASDLRKHE--RTHPRVPMGPTPLEPIVALLGMP 493
OY 983 GCSPS 987
DB 494 EEGPA 498
RESULT 33
AAB95116
ID AAB95116 standard; Protein; 725 AA.
XX
XX AAB95116;
AC
XX 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17104.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 17104; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632


```

Db      61 MSLEKSSDDPSVPQNPPTP---LGHSN-PL-----DHQIPD----- 93
Oy      210 SLGQTVGAPASPSLP-----GTGTASTKPLPLFSPDKPYQTSKTLAS 254
Db      94 -----PPAPEVVPPTPSDWTKACASMQMGALTTWNSPPVVP-----AN 131
Oy      255 SSSSSSSSGAATPKQAFFHLNPLPSQHPFSAGGVSRHKPTAPSPALPGSTDLIAS 314
Db      132 EPSLRELVOG-----RPAGAEKPYICNECKS----- 158
Oy      315 PHLAEPSTGTLAAGCIGAARGLEATASPGLLKPKNGSGELSYGEVMPLKPGGRHKCR 374
Db      159 -----FSQMSKLLRHQRIHGERPNTSECG--KSFQSSHLVQHQRTTGERP---YKCP 209
Oy      375 FCAYVFGSDALQIHLSRHTGERPYKCNVCNRPRTTGNLKHVHRHREKYP----- 426
Db      210 DCGKEFSMSSNLVQHQRTTGERPYKCTECEIAFTQSTNLKHQSRHTGERPYKCGECRR 269
Oy      427 -----HYQMNPHVPPEHLDYITSSGLPYGMVPPPKAEFEATPGGVER 472
Db      270 AFYRSSDLIQHATHGERPKYKCP-----CGKRGQHNLKHKQKHA-----GE 315
Oy      473 KPLVASTALSTESTLTS---TSAGTATAPGLPAFNKF---VLKMAVEPKNKADENT 525
Db      316 KPYRCTEGCKSFQSSSELTOHQRTHTGERPECELECGKSFQHSSTLKHQRTHLREDPER 375
Oy      526 PGSEGSALSGVAESSTATLMQLSKLMTSLPSMALLTNHFKSTGSPF--LPLCARALGAS 583
Db      376 CP-----VCGKTFTLATLLRHQR-----THTGERPYKCPKC-----GKS 410
Oy      584 PSETSKL--QQLVERKIDRCQAVAVTSAASGAPTSAPARSSASSGPNQCVICLRVLSCP 641
Db      411 FSVSSNLNHRHQRHGER-----PYCACGCKSFIMS 442
Oy      642 RALRLHYGQHGGERPKCKYCGRAFTRGNLRAHFVGHKASPARAQNQSCPICQKFTNA 701
Db      443 STLIRHQRIHTGKPKYKSCGKSFIR---SSHLIQRIHTHTGERPKYKCEGCKSPFSQS 498
Oy      702 VTLOOHVRNHLGQI---PNGGIALPREGGAAQENGSEOSTVSGAGSPQOQSQDPSPPE 758
Db      499 SNLITVHTHMDNLEFVCSGCKAFLEAHELEQHR-----VIHERKKTPARRAO----- 547
Oy      759 ELSEEEDEDEEEDVTDBDSIAGRSES-----GGEK---AISVRDSEASAGAE 809
Db      548 -----GDSLGLGDDPSLITPPGAKPKKCLVCGKGFNDGCTFMQHQ 588
Oy      810 -----VGTVAATAAGKEMDSNEKTTQOSSLPPPPPPSLDOPQPMEOGSSGV 857
Db      589 RIHIGENPKYKNAAGLTAHNA-----PKPQLRSPRLPFRGNS-- 625
Oy      858 LCGKEGCKPERSSSPASAL--TPREGATSVTLVEFLSLQEMARKKEGESSRK--ACVCG 915
Db      626 YPGAAAG---RAEAPQOLKPPREG-----OEGFSQIRGLLSSKITYICHG 668
Oy      916 QAPPSQAALEEHOKTHPEKGPL-----FTCYFCROGFLERAT 952
Db      669 ESTLDSVLLQHQITHTGNEKFFLPDVRIGLGEAGAPSPFLSGPKPKCECKOSF-----G 724
Oy      953 LKKHMLLAHHQV 964
Db      725 LSESELL-HQKV 735

RESULT 35
ABG06465
ID      ABC06465 standard; Protein: 882 AA.
XX
AC      ABG06465;
XX
DT      13-FEB-2002 (first entry)
XX
DE      Novel human diagnostic protein #6456.
```

```

XX      Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW      food supplement; medical imaging; diagnostic; genetic disorder.
XX      Homo sapiens.
OS
XX      WO200175067-A2.
XX      11-OCT-2001.
XX      30-MAR-2001; 2001WO-US08631.
XX      31-MAR-2000; 2000US-0540217.
XX      23-AUG-2000; 2000US-0649167.
XX      (HXXE-) HXSEQ INC.
XX      Drmanac RT, Liu C, Tang YT;
XX      WPI: 2001-639362/73.
XX      N-PSDB: AAS70652.
XX      New isolated polynucleotide and encoded polypeptides, useful in
XX      diagnostics, forensics, gene mapping, identification of mutations
XX      responsible for genetic disorders or other traits and to assess
XX      biodiversity.
XX      Claim 20; SEQ ID No 36824; 103pp; English.
XX      The invention relates to isolated polynucleotide (I) and
XX      polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX      and gene mapping, and in recombinant production of (II). The
XX      polynucleotides are also used in diagnostics as expressed sequence tags
XX      for identifying expressed genes. (I) is useful in gene therapy techniques
XX      to restore normal activity of (II) or to treat disease states involving
XX      (II). (II) is useful for generating antibodies against it, detecting or
XX      quantitating a polypeptide in tissue, as molecular weight markers and as
XX      a food supplement (II) and its binding partners are useful in medical
XX      imaging of sites expressing (II). (I) and (II) are useful for treating
XX      disorders involving aberrant protein expression or biological activity.
XX      The polypeptide and polynucleotide sequences have applications in
XX      diagnostics, forensics, gene mapping, identification of mutations
XX      responsible for genetic disorders or other traits to assess biodiversity
XX      and to produce other types of data and products dependent on DNA and
XX      amino acid sequences. ABG00010-ABG30377 represent novel human
XX      CC diagnostic amino acid sequences of the invention.
XX      CC Note: The sequence data for this patent did not appear in the printed
XX      CC specification, but was obtained in electronic format directly from WIPO
XX      CC at ftp.wipo.int/pub/published_pcl_sequences.
XX      SQ      Sequence      882 AA;
XX
XX      Query Match      5.9%; Score 311.5; DB 22; Length 882;
XX      Best local similarity 21.9%; Pred. NO. 1.1e-09;
XX      Matches 180; Conservative 82; Mismatches 304; Indels 255; Gaps 34;
Oy      18 PABLGDSASEBDHQVCAKCCAGFTDPTEFLAHQNCSTDPVWVITGGENPNSSASS 77
Db      129 PLQVLSGSHKEKENLKP-AKAOLPFTTTPGFTTNDREKESP-----SSTWNPSEMQRKA 181
Oy      78 EPRPEGHN-----NPQVADTEHSNPPDSGSSVPTDPTWGPERRGESSGHFLVAAT 128
Db      182 SPRKWKHHNQRPSTHKVKKMWSSEOMKLPSTKKAEP--PTWQOLKK----- 226
Oy      129 GTAAGGGGGLLSPKLCATPLRPESTPAPPPPPPPPGVSGHLNPLILEKYLQ 188
Db      227 -----LTOLPKKS-----LENTKV-- 240
Oy      189 QROTHQMQMTQICROVLLLSLGQTVGAPASPSLPSTGTASTKPLPLFSPI-----K 244
Db      241 -----TOTPENTLIAALMIVS---TAGAAAA-----NTTWAYVYVPPPLIRAVTMDN 285
```

```
OY 245 PVQGTSKLASSSSSSSSGAEPTPKQAFHLYHPLGSOHPFSAGVGRSHKPTPAPSPAL 304
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 286 PIEVYVNNNSAWMPGPTDDRCPRDKPEEGEMMINISTGYRYPIC--LGR-----APACLM 337
OY 305 PGSDQLLASPHLA----PPSTIGLLAA-----OCLGAA-----RGLSAT 341
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 338 PAIOMWLEVPVPGTTSRTPTTYHTGLTFRDVAIEFSQEWKCLDPALEDSTYRRKANSQ 397
OY 342 SP-----GLTKPKNG-----SGE-----LSYGEVWGLPEKPCG 369
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 398 KPYCNCEGKKAFTQNSNLISHRHISGEKPYKCECGKFTYVRNSLTIHQYIHTGEKP-- 455
OY 370 RHKRFPKAVFGSDSALOIHLSHTEGERPYKCNVCGNRFTRGNLKVAFHHRREKYP-- 426
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 456 -YKCHECGKVFRRHNSYLATHRRRIHTGEKPYKCEGKAFRQHSNLTTHQLIHTGKPFKC 514
OY 427 -----HYQMNPRVRENHLYVITSSGLPYGMSV-----PREKAEF-- 461
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 515 NECKKLTQNSHLISHWRITHGEKPYKCEGKAFSVRSSLAIHQTIHTGEKPYKCEG 574
OY 462 EAATPGGVERKPLVASTATLSATESLTLTSTAGTATAPGLPAFNKFLMKAVPKNKA 521
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 575 KVFVGRGV-----IGTCLLSVFQYIVTSSRS-----RASPNRA 612
OY 522 DEN-----TPPSGSAISGVASSTATLMQLSKLTMSLPMSMALLTNHFKSTGSP--LPL 575
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 613 TQSLILAPPTSRGSG-PGPA-SVLRRLAOTRRKM---AW---TESCTAACAFPSCLVL 663
OY 576 CARALGASPSR---TSKLOQLVEKIDROGAVAVTSASGAPTSAPAPSSASSG----- 627
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 664 LYRRGV-PTLOQVPAPLFSTOSKLRPSRTKSSLAHORIHTEGKPYKCEGKVPFSQ 722
OY 628 -----PNOCVICLRYLSCPRALRLHYGOHGERPFKVCGRAFSTRGNL 672
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 723 TSSLARHWRITHGEKPYKCEGKVFYSNLSHSHRVHTGEKPYKCEGKAFSVHNL 782
OY 673 RAHFVGHKASPAARQNSCPICQKFTNAVTIQQHYRMHLG 713
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 783 TTHOVITHGEKPYK---CNCGKGFVSHSILTHOVITHG 819
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 36
ABR69958
ID ABR69958 standard; Protein: 934 AA.
XX
AC ABR69958;
DT
XX
DE 26-MAR-2002 (first entry)
XX
KW Drosophila melanogaster polypeptide SEQ ID NO 36666.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN MO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001MO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
PI
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
XX
XX N-PSDB; ABL14061.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
```

```
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX PS Disclosure: SEQ ID NO 36666; 21bp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pot_sequences.
XX
XX SQ Sequence 934 AA;
XX
XX Query Match 5.9%; Score 311.5; DB 22; Length 934;
XX Best Local Similarity 18.6%; Pred. NO.1.2e-09;
XX Matches 195; Conservative 91; Mismatches 387; Indels 373; Gaps 34;
XX
OY 48 LAHONACSTDPPVM-----VIIGQENPNNSASSSEPRRGGHNNPQVMDTEHSNP 97
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1 MQHVAASAVSPSVTPVPTTGTTILGGPRPLKSEHKEDGR-----PHGIEMKYNI 55
OY 98 PDSGSGVPTDPWGPERRKEESSGHFLVAATGTAAGCGGCLILASPKIGANLPPESTPA 157
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 56 EDISQLFTYHEVEG-----KIHGDVYVNHQLAANGGL----- 88
OY 158 PPPPPPPPPGVSGHNLIPILFELRLVLDQRIHQOMQTEQICRQVILILGSGQTGVA 217
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 89 ---PPRPPLPQVYS-HAASAAAAAASSTNNAVAANVAASANAAAAAASAG--GL 142
OY 218 PASPELPGTGAASSTKPLLPSPIKPVQTSKTLASSSSSSSSSGAEPTPKQAFHLYH 277
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 143 PPATS---GNCG-----QQVTVTTSSTSGSGTSGTGTAGELMPK 185
OY 278 PLGSOHPESA---GGVGRSHKTPAP--SPALPGS-----TDQLASPHLAF 320
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 186 MEGSIHGVDGSGNGGQGNVALAPDGPATGTCIDCGMKFQRYQLTV--HRRYH 243
OY 321 STTGILAAQCLGAARGLATASPGILKPKNGSGELSYGEVMP-----L 364
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 244 SERKRFMQVCGGQ---FTTSQDLTR---HGKIHIG--GFMFTGICVNNVANNSTL 292
OY 365 EKPGGRHK-----CRPKAVFGSDSALOIHLSHTEGERPYKCNVCGNRFTRGNLKVAF 418
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 293 ERHMKRHSIDKRFACCTICQKTFARKENHLDNHFSHTGEPFRCQYCAKFTTRKDHVNHV 352
OY 419 HRHREKYPH---YQMNPRVRENH--DYVITSSGLPYGMSVPREKAEELATPGCGVERK 473
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 353 RKHTGEPHRCIDICKKSTTRKHYYNHHMTGQTHQCDVCGKK-----YTRK 401
OY 474 PLVASTATLSATESLTLTSTAGTATAPGLPAFNKFLMKAVPKNADENTPPSGESA 533
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 402 EHLANHRSHTNET-----PFRCEICGKSFSRKEN----- 431
OY 534 ISGVASSTATILMQLSKLTMSLPWALLTNHK-STGSP--LPLCARALGASPSSESKL 590
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 432 -----FTNHLIHTGEPRHRCDFCSKTP-----TRK 457
OY 591 QQLVERKIDROGAVAVTSASGAPTSAPAPSSASSGPNOCVLCRLVSCPRALRLHYGQ 650
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 458 EHLINHY-RQ-----HTGESPRHCSTCKTFTYRKENHLYVNIHQ 494
OY 651 HGERPRPKCVCGRAFTSRGNLRAHFVGHKASPAARQNSCPICQKFTNAVTIQQHYRM 710
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 495 HTGETTPKCTYCTKAFPRKD---HMVNHVROHTGSGPKCTYCTTFTTKENHNTNVL 550
OY 711 HLGQQLPNGGTALPREGGAOENSGOSTVSGAGSPDQOOSQSPSEDEELSEEEDEE 770
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

Db 551 HTG-----DSPHKCEY 561
QY 771 EEEVVTDEDSLAGRSESGEKAISVRGDSSEASGAEEVGYVAAATAAGKEMDSNETT 830
Db 562 COKETFRKEHL-----NNHMR 577
QY 831 IQSSLPPEPPPPDSLDQPPMEGSSGVLGKKEGCKPERSSSPASALTPEGEATSVTLVE 890
Db 578 QHSS-DNPHCCNVCKNPKFTTRKHLINHMNRCHTGRD-----FTCEICGKSFPLKG 627
QY 891 ELSIQEAMRKPERSESSSRKACEVCGOAFPSQALAEHQTHPKRGPLTCVFCRQGFLE 950
Db 628 NLFPHQSHSTKGEEMEREFACCKPKNFICKGLVSHMRSHSGEKP--HACTYLCSKAFVER 686
QY 951 ATLKKHMLAH-----HVOEPFA 968
Db 687 GNLRHKMKMHPDAMPPRPVHPHPIAGVLTQVKQEVKPIILPHHSATTTMTHTIQOIT 746
QY 969 PHGPNIMALSLVPGCSPTSTGSL 994
Db 747 AGAAGGAGAVQLTPEGLVPLVTSTLIS 772
RESULT 37
ID ABB59829
AC ABB59829 standard; Protein; 1893 AA.
NC ABB59829;
CD 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 6279.
DS
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
OS Drosophila melanogaster.
PN HQ200171042-A2.
PD 27-SEP-2001.
PF 23-MAR-2001; 2001WO-US0231.
PR 23-MAR-2000; 2000US-191637P.
PS 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PMD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL03932.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure: SEQ ID NO 6279; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX Sequence 1893 AA;
SQ

Query Match 5.8%; Score 308.5; DB 22; Length 1893;
Best Local Similarity 20.1%; Pred. No. 4,1e-09;
Matches 245; Conservative 132; Mismatches 428; Indels 413; Gaps 55;
QY 53 ACSTDPVAVIIGQENPNSSASSEPRRPGHN-NPQVMDTEHNPDDSGSVPTDPFWG 111
Db 397 SCSTD-----VSSGSHSRSSSSSLNNNNNNHKAANNMLKLE-----ELEVESTEDQDEN 446
QY 112 PERGEESGHLV-----AATGTAAGCGGGLLSPKLGATPLPPES 154
Db 447 KORRLKTTINNMIIESEQEDMDDEADADVAALMTSTPPDVATLLAGASAGA-----A 500
QY 155 TPAPPPPPPPPP-----PGVSG-----HLNPLILELRYL-----CQR 190
Db 501 SRSPTPSPASALLISCACGASDEFTLRALCVHLDAMSHDIPRAKCDCEVIFATHRL 560
QY 191 QIHOMQMTQICR-QVLLGSLQTVGAPASP--SELPGTASTKRLPLPSPIKRVQ 247
Db 561 QSH-----CCRLPVALAGLPPLLGASSSPLNHEPEDEHGDDEL-----E 603
QY 248 TSKTLASSSSSSSSSGAETPKQAFHLVHPLSGQHPPSAGVGRSHKPTP----- 298
Db 604 QKERLASOSED-----FFHQLYLKHTKATNCAGALSHPSPIKHEPADTK 647
QY 299 -----ASPALPGSTQDLIASPHLAFPTTGILAAQCLGAARGLATASPGL- 345
Db 648 DLADIQSLNMTSSSSSFLRNFQSVTPNNSQVSLDS-----RDQEEHQDAPT 697
QY 346 -----LKPNSGELSVEGYMPLKPGRRHKCRCAKVFSS 382
Db 698 SEFRRMKILRGEFPCKLCTAVFPNLRLALKGNRVHLGAV-----GPAQGFRCMWCYPAYCD 752
QY 383 DSALQIHLSHTGERPYKVCVCGNRTTTCGLKYHF--RRH-----REK----- 424
Db 753 KAALVHMRTHNGDRPECAVNAFTTKANCERHLNRHNGKTSSEVKRAIYVHPAEDA 812
QY 425 -----YPIVQMNHPVP-----EHLDYVITSSGLPYGMSVPERKA 459
Db 813 GGEDSKSRIGEDLADISFRISPTPPRPVNESKQLKHM--LQENHLAPVNOQPPRLKI 870
QY 460 E-----EEAATPGGVERKPLVASTTALSATESLTLLSTAGTATAPGLPAFNK 508
Db 871 QVKSILQDLVDKRPAPAPQOQOQOQOQOQESGALDFSMVDLSKRK--TGASILTPAVTR 929
QY 509 FYLMKAVPERKNKADENTPPSESGAISGVAAESTATLWQ-----LSKLTMTLPSW 558
Db 930 TPTPAVAVPY-----TPGGVGTPTDLAAAIQOQQLLAQOQLFGAGGEYMOQLFRSL--- 980
QY 559 ALLTNHFKS-TGSEFP-LPLCARALGASPETSLSKLOLVEKIDROGAVAVTSAAGAPTSS 616
Db 981 -----MFQSGTSGFPPFPFPA-----PPPOQANPEKPPMVSPPRINMPV-GVGVGPV- 1029
QY 617 APAPSSASSGPNOCVYICLRLVLSCPRALRLHYGHGGERPEFKCYCGRAPSTRGNLRAHF 676
Db 1030 -----PPGPAVMYIKNGVL-MPKOKORYRT---EPPFACHCOSARFTLRSNMERHV 1078
QY 677 -----VGH-----KASPARA----- 687
Db 1079 KOQHPOFYAQRORSRSHVMMRGASNVAAAAAATAATVMAGPGSGSGFNHNGH 1138
QY 688 -----QNSCPIQCKKFTNNAVTLQOHVRMH-----LGQILPNGGTL-----PE 725
Db 1139 GHGSHSHGHAPISEQ--VKCALIAQOLKAHKNTDLLQOALAHGSSVAGNPLLRGYPL 1196
QY 726 GGGAAOENGSEOSTVSGASFPQOQOQPS-----PEEELSEEEDEDEEE 771
Db 1197 TNPSPMHNSSQ-----GNGQATAMDDDEPKULIIDEDENHDEVEEDVDDEDEDEEE 1252
QY 772 EEEVVTDEDSL-----AGRSESGE--KAISVRGDSSEASGAEEVGYVAAATAAGKEM 823
Db 1253 MDEPEDEPELILDEQPAKEAEEDEQLPLEQLDTKEAQAQMAETI--LEQAIKAGKPL 1310

QY	824	DSMEKTTQSSLPPEPPDSDLPD----	QPMQGS--	CVLGKKEGGK--	PERSSP----	873
Db	1311	-SPPPEKSNSPANPTVATTWQEP	AITAPSTNPSLKTMTIAQAEV	EGSLKEVASSPFXD		1369
QY	874	-ASALTPPEGE----	ATSVT-----	-----	-----	896
Db	1370	ESQDLVPVALVDNATNSQNN	GFNSYFRPSVDANHMEQSD	EGIVASSGSASESNNGTE	DV	1429
QY	897	---AMKREPESS-----	SRKACEVCGQAFPSQAAL	EEHOKTHPKESGLEFTVCF	ROGF	947
Db	1430	TSSSSSEPEPKKSAVSLAP	NRVSCPYCQRMFPWSSSL	RRHILHTHGQKP-FKCSHC	PLIF	1488
QY	948	LERATLTKHMLAH	HOYQ	965		
Db	1489	TTKSNCDRHILRRKH	GNVE	1506		
RESULT 38						
AAM39029						
ID	AAM39029	standard; Protein; 722	AA.			
AC	AAM39029;					
XX	22-OCT-2001	(first entry)				
DE	Human polypeptide SEQ	ID NO 2174.				
KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;					
KW	peripheral nervous system; neuropathy; central nervous system; CNS;					
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;					
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;					
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;					
OS	Homo sapiens.					
PN	WO200153312-A1.					
PD	26-JUL-2001.					
PF	26-DEC-2000; 2000WO-US34263.					
PR	21-JAN-2000; 2000US-0488725.					
PR	25-APR-2000; 2000US-0552317.					
PR	09-JUL-2000; 2000US-0598042.					
PR	19-JUL-2000; 2000US-0620312.					
PR	03-AUG-2000; 2000US-0653450.					
PR	14-SEP-2000; 2000US-0662191.					
PR	19-OCT-2000; 2000US-0693036.					
PR	29-NOV-2000; 2000US-0727344.					
XX						
PA	(HYSE-) HYSEQ INC.					
PI	Tang YN, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;					
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;					
PI	Zhao QH, Zhou P, Goodrich R, Demanac RT;					
XX						
XX	WPI: 2001-442253/47.					
DR	N-PSDB; AAI58185.					
PT						
PT	Novel nucleic acids and polypeptides, useful for treating disorders					
XX	such as central nervous system injuries -					
PS	Example 4; SEQ ID NO 2174; 10078bp; English.					

CC	lateral sclerosis, and Shy-Dreager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	activation/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
CC	
XX	Sequence 722 AA;
XX	
Query Match	5.7%; Score 303; DB 22; Length 722;
Best Local Similarity	22.4%; Pred. No. 2,6e-09;
Matches 164; Conservative 70; Mismatches 271; Indels 228; Gaps 27;	
QY	348 PKNKGSGEISYGEVWGPFLKPGGRH-KCRFCAYFGSDSLQIHLRSHGTGERPYKNCN 406
DB	33 PAYSAGSLGMGAVMSSESERAGERRRPCPCGRFRNRSLIALHLRAIPGAGAPCCPHGH 92
QY	407 RPTTRGMLKAFHNRKREYPRVQAMPHPVRPHLDVITYTSSGLPYGMSVPPKAEAEATP 466
DB	93 RAAQALLRSLRTHQRP
QY	467 GGGVERPPLVASTALSTESLTLLSTAGTAPGLPAFNKFLVLMKAVBPKNKADENTP 526
DB	113 ----RSP--AARLLLEERALLREARLGRARSSG-----GQAT 146
QY	527 PGSEGSASISGVAESSTATLMOLSKLMTSLPSWALLTNHFKSTGSFPLPLCARLAGASP 586
DB	147 PATEGLARPOAPSS-----AFRCPPCKGFRFS-AE 177
QY	587 TSKLOQLEVERKIDPGAVAVTSA-----ASGAP-----TTSAPAPSSASSGPN 629
DB	178 RERHNLTLNHRKWKCGLCSFGSSQEBELHLSLTAGAPERPLAATSAAPROPOROPPO 237
QY	630 QCVICLVLSCLRALRLHYGQNGER-----PRCKVCGRAFTSGN 671
DB	238 ----PEPRVPOPEPEPEREATPTPAPAPAEPPAPPEPRQCVCQSFTQSWF 288
QY	672 LRAHFVGHKASPARAONSCPIQCKKFTNNAVTLQOHVRNH-----LGGIIPNGTAL-- 723
DB	289 LKGMKRNKHS----FDHACPVCGRCFEPWFLKNHMKHASKLGLPARGPASGPAPAP 344
QY	724 --PEGGAQENGSEOSTVSGAGSFPOQSOOPSEBELSEEBEEDDEEBEDVTEDSL 781
DB	345 QRPDLGLLAEYELGRLALLAPAT-PAERREPRSLILGYLSLRAGEGR-----NGEGAE 397
QY	782 AGNGSESG----EKALSVGRDSEASAAEERYGVYAAA---TACKEMDSNEKTTQSS 834
DB	398 PGGRSGFSGRPLSSALPARARRHRAEPEEEVEAEQETWARSRSLG-----LAS 451
QY	835 LPPRPDPSIDOPRPMQSGSVLGGKEEGGKPERSSPSALATPREGATSVTLVEELS- 893
DB	452 LHRPR-----GGGRGHSASAGA-----QARSTATQEEINGL 482
QY	894 LQEAHMKKEPESSSRKACEVCGAPPPSOALAEHOKTHPKESPLTFCVPCROGFLEBATL 953
DB	483 LVGGTRPEEGRGATGKCPCFGSKFSRANHNLKVHLRVHNGEER-XYCPHCDYAGTOSGL 541
QY	954 KKHMLLALHNOV-----PFAP-----HGPNITALSLVP-----GCASPSTSTG 992
DB	542 KHN-LDRNHREQSGAGRGPRPPPPPSQKSAPQSGAKPSPQATWVGASSPAPPSSG 600
QY	993 LSPFPKDDPTIP 1005
DB	601 AGPGSRR-KPASP 612
RESULT 39	
ABB59449	
ABBS9449 standard; Protein; 744 AA.	
ABB59449;	

XX 26-MAR-2002 (first entry)
DT Drosophila melanogaster polypeptide SEQ ID NO 5139.
DE Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
KW Drosophila melanogaster.
XX WO200171042-A2.
PN 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
XX MPI; 2001-656860/75.
DR N-PSDB; ABL03552.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -
PS Disclosure; SEQ ID NO 5139; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL01840-ABL6175), expressed DNA
sequences (ABL01840-ABL6175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/Pub/Unpublished_pct_sequences.
XX
SQ Sequence 744 AA:

Query Match 5.7%; Score 298.5; DB 22; Length 744;
Best Local Similarity 17.1%; Pred. No. 4.9e-09;
Matches 174; Conservative 88; Mismatches 325; Indels 429; Gaps 24;

QY 90 MDTESNPDDSGSVPTDPTWGPERRGEESGHLVATGTAAGGGGILLASPKG--- 146
DB 1 MPTSSSSSEITSGGGGAIPLRLPSRMDQFMNS---MAAAAAAGGGGLPGADRNGGSG 56

QY 147 -----APPPLPESTRP-----APPPPPPP----- 164
DB 57 GSDGSGONGNGSRNSASRIASVENQLAYQHLAGLHGPRPPRRSHHREISAFVPLV 116
QY 165 ---PPPPGVSGHLNPLLELFAVLQOQIHOQMTEDQICRQVLLGSGIVGAPASIS 222
DB 117 TGVKVRPGSNSNYEIIIMAMDREKRELALRE-----AAAAAAMGRRPGGCGGPG 164
QY 223 ELPGGTASTRPLPLFLSPIKIVQTSKTLASSSSSSSSSSSAETPRKQAFHLYHPLGSG 282
DB 165 -VPPPGV-----LYGAGVP 178
QY 283 HFFSAGVGRSHKPTAPSPALPGSTDLIASPHLAPPSTGLAACL--GAARGLAET 340
DB 179 PEPVLTG-----GSPSTGAGS-----FPFPGAAAAALPRPGIGPMHAG 219
QY 341 ASPLGLKPKNGSGELSYGEVGLPEKPGGRHKRCRCACAVFGSDSLQIHLKRSHTGERPK 400
DB 220 LDRRLLR-----APGRASRPKKOFICKFCNROFTKSYNLLIHERHTDERPYS 267

QY 401 CNVCSNRFETTRGNLKVHFHRHREKYPHVQMNPHVPREHLDVITSSGLPYGNSVPERKAE 460
DB 268 CDICKAFRRQDHLDRKHRIHSKEP----- 293
QY 461 EEATPPGGVERKPLVASTTALSATESLTLSTAGTATAPGLPAENFKVLMAKAVEPKK 520
DB 294 ----- 293
QY 521 ADENTPPGSEGAISGVAESSTATIMQSLKMTSLSPSMALLTNHFKSGSPPLPCARAL 580
DB 294 -----FK----- 295
QY 581 GASPESTSKLQQLVEKIDROGAVAVTSAAGAPTTSA PAPSSASSGPNOCVICYLRLSC 640
DB 296 -----CTECKGFGCQ 305
QY 641 PRLRLHYGQHGGERPFKCVCGRASTRGNLRAHFYGNKASPAARAONSCPIQCKKFTN 700
DB 306 SRTLAVNHKILHMEESPDKCPVCSRFNQRSLNLTHTLTH---FDHKPEYECSSCGKVFRR 361
QY 701 AVTLQOHVAMHLGGQILPNGSTA-----LPEGGG 728
DB 362 NCDLRNHALTTHAVGEVNSGDYVDVEEDEARNLSGDEEDSLLEVDSPROSPVHNLGECSG 421
QY 729 AAOENGSEOSTVSGAGSFPOQSOQP-----SPEELSE-----EEEDEDEEDYU 776
DB 422 SEKSESEMRKLKKAALIDHEESEFDDDEDELDLPVHNDLPREDDDFRDEEDQ 481
QY 777 DEDSLAGR-----GSESGEKAISVRGDSSEASGAEEVSTVAANAATGKEMDSNEKTT 830
DB 482 AEVALVARQASKAANTSOSSSVGTKEPREGVYHCHNEGGETYIMRPHGEKHQEPGNS 541
QY 831 QOSSLP RP-----PPBSLDQRPMEQSGSVLGKDEBGKPRRSSPA--SALTPEG 881
DB 542 GIASLPVPSEFVRYSVPPGAAGRP-----APGAPRTNQHGNHILP 589
QY 882 EATSVTLVELSLQEAMRKPKGESSSRKACSEVCGAEPGQAALBEHOKTNPKRGRLTCV 941
DB 590 D-----PYRLIHYNRDLHNKSLNSKAGVPRPHRPTIIQPESGKPRPQRLHSPH 642
QY 942 FCRGGLERATLKKHMLANHOVQFRAPIHGFONTAALSILVPGCSPTISTGSLPFP 997
DB 643 EAMPSEFGSIPMKRILPA--PTIDLMDPHNHGIGQRTFVD--SPSIVALNMSRNP 695

RESULT 40
ABBB6002
ID ABB66002 standard; Protein; 744 AA.
XX
AC ABB66002;
XX
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 24798.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;

ALIGNMENTS

RESULT 1

US-09-041-886-23

; Sequence 23, Application US/09041886

Patent No. 6235872

GENERAL INFORMATION:

APPLICANT: Bredesen, Dale E.

APPLICANT: Rabizadeh, Sharroz

1. TITLE OF INVENTION: Proapoptotic Peptides, Dependence

; TITLE OF INVENTION: POL

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell & Flores LLP

STREET: 4370 La J

CITY: San Diego

STATE: California

COUNTRY: U

COUNTRY: United States
ZIP: 92122

```

'
' ZIP: 92122
'
' COMPUTER READABLE FORM:
'

```

COMPIER READABLE FORM:

MEDIUM TYPE: FLOPPY disk

COMPUTER: IBM PC compatible
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentTr Release #1.0

SOFTWARE: 1

CURRENT APPLICATION

APPLICATION NUMBER: US/C
FILING DATE:

FILED DATE:
CLASSIFICATION:

CLASSIFICATION:
ATTORNEY / AGENT INFORMATION:

ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Gerald

NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31 015

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ I

SEQUENCE CHARACTERISTICS

LENGTH: 1185 amino a

TYPE: am
;

US-09-041-886-23

9
0
6
7
8
9
C
C
C
B
C

```
Query Match 5.5%: Score 289.5: DB 4: Length 1185.
```

```

Best Local Similarity 21.5%; Pred. No. 4,3e-11;
Matches 231; Conservative 95; Mismatches 361; Indels 387; Gaps

```

2y	55	STDPPVMTIIGGOENPNSSASSEPRE---	GHNPNQVMDTEHSNPDGSSVP-----	105
2b	168	SPQPP-----DSTRPEASEEPHPSTVPTG	GHAPMEPPTSRMFOAPGAPPPHPQY	220
2y	106	-----TDTPMGPERGEESEGHFLVATGTA	AGGGG-----LLASFKLGA	147
2b	221	PGGTGVLSPFMGRGGGAAS-----	VGGPNGQKHPPPTPIISVSSGASGA	270
2y	148	TPLPPESTPP-----APP-----	PPPGVGS-----GHIN	176
2b	271	PPTRPPTPVVGGNLPASAPPAPFPHTPML	PPPLRLNNASAPGLAQPLDGHLP	330
2y	177	IPLLLELRVLQQRQLHQOMTEQICQVLLG	SLGCTVG-----APASPEL	224
2b	331	SPY-----AMGGMGGLPGEPEKPTLAP-	SPHSL	359
2y	225	PGTGASTKPLPLPSPILKPVQTSKTLAS	SSSSSSSSSGAETPKQ-----APFHLYHPG	280
2b	360	P---PASSAAAPAKRFVYSSSSSSAAAS	SSSSSSSSSAPFAQAQLPSTPHSPFP---	414
2y	281	SOHPFSAGVGRSHKPTPAPSPALPSTDL	ILASPHILAFSTTGILLAAQCTGANGLEAT	340
2b	415	---PTS---LSVSNQPKKTOPSLP---	SOAVNSQGGPPPPPYGRLLA-----NSN	456

```

OY 341 ABGILKPKNSGSLSTGVCMGRLEKRGSKAKRCSCAKVFGSDSLQILNRSTGER -- 397
Db 457 AHGRRPSTGAOSTAH---PRVSTNNHNOOQ-----0000000000HNGSGRP 505
OY 398 -----PYKWCWCGNRFTTNGMLKYNHNRHREKUPHYOMNHNVC-RENH---DUYTTSSG 447
Db 506 PRGAFPH-----PLEGSSHNHNHYAMSPSLGSLRPPRRGRANLRRPHNSQVYSQA 556
OY 448 LPRGMSTPRPEKAEEAAT-----PGGVEKRPVASTTALSTATES-- 487
Db 557 GRPG---PRVSSSSSSSSSTSGSYRCSNRPSSQGRPYRFPRTYTTSSATLSTVL 613
OY 488 LTLTSTASG---TATAPGRAPNK-----FVLKAVEPKKKADE-----NTPRGSGS-- 532
Db 614 ATYASSSAGKKTASPPRRPPRYGKAPRSBGAKYKTALPRGKTPGSSPSFRGTGPRYGTSP 673
OY 533 -AISVAESSTATLMQSLMTSLRSMWALTNNHFKSTGSPRLCARALGASPSERKLU 591
Db 674 PARGTFPKPSSPIVGP--GRLPRPAPS-----GLPSLPPRPAAPASGRLSANQINO 723
OY 592 OLVEKIDROCAVAVTSAASCAPTTSAPARSSASGPRMOVCYCLAVLSCPRALRLHYOH 651
Db 724 EPRAEET-----PRESPPRARSBPP--KVVDVP-----SH 754
OY 652 GGERPRCKVCYCGAFSTNRGNLRHFVGHKAPAPAAQNSCPIQCKFTNAVTLQOCHVNM 711
Db 755 ASQSAFPHKHLDRGFCNSCARSDLYFVPLEGSKLAKR-----ADLVEKVRRE 801
OY 712 LGGQITNGGIALPEGGGAQDENSEQSTVSGASFPOQSOQSPREBELSEEEDEDEE 771
Db 802 -----AEGRAAREKEREREREKERERER 826
OY 772 EBDVYDEDSLGRGS---ESGGEKALISVRGDSSEASGAEEEGVAAATATACKEMDSNE- 827
Db 827 ERELRESVKLAQGRAPVECRSLGRVPHNRPEFGS-----AAATVPYLCGRDPR 876
OY 828 ---KTIOQSSLPRPPDSDLDP--QRPEDGSSGYVLGKEBEGKPERSSPASALTPEGGA 883
Db 877 ALRTLSEYARPHVMSGNRNMHPYUULGAVDPGLG---YNVALYUSSDPAAREREREA 932
OY 884 TSVTLVEELSLDEAMKREPRESSSRKACEVGOAGFARPSOALBLEONKTHPKBEPLTTCVC 943
Db 933 -----REMDLRN--RLKRP-----FEVKPS-----ELEPLHGVCRP----- 961
OY 944 RQGLERATLKKHMLLHANNOVOPFARPHGRONIALSLVPGCSPTITSGLSPPR 997
Db 962 -----GLDPFRPG---GLALDPG-----PGLHFPFR 985

```

RESULT 2
 US-08-475-844-9
 : Sequence 9, Application US/08475844
 : Patent No. 5972643
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Lobanenkoy, Victor V.
 : APPLICANT: Neiman, Paul E.
 : APPLICANT: Klenova, Elena M.
 : APPLICANT: Goodwin, Graham H.
 : APPLICANT: Filippova, Galina N.
 : APPLICANT: Collins, Steven J.
 :
 : TITLE OF INVENTION: CTCF
 :
 : NUMBER OF SEQUENCES: 21
 :
 : CORRESPONDENCE ADDRESS:
 :
 : STREET: One Market Plaza
 : CITY: San Francisco
 : STATE: CA
 :
 : COUNTRY: USA
 : ZIP: 94105
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible

Db 710 DABNGDLTPE 719

RESULT 4

US-09-362-123A-4
; Sequence 4, Application US/09362123A
; Patent No. 6451558
; GENERAL INFORMATION:
; APPLICANT: Cooke, Michael Paul
; APPLICANT: Holmes, Claire
; APPLICANT: Strenko, Oksana
; TITLE OF INVENTION: No. 6451558el Genes in the Control of Hematopoiesis
; FILE REFERENCE: 4-30629A/SYS
; CURRENT APPLICATION NUMBER: US/09/362,123A
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: 09/128,310
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 543
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-362-123A-4

Query Match 4.7%; Score 246; DB 4; Length 543;
Best Local Similarity 20.7%; Pred. No. 1.2e-08;

Matches 143; Conservative 54; Mismatches 217; Indels 278; Gaps 22;

QY 136 GGLILASPKLGATPLP-----PESTRAPPPPPPPPPGVSGHLMPLILEELRVL 187
Db 53 GLLLTMMPOATTVSLPGSWSLTEDRDVSGEMPRAPRPTPGMTTSVFPVAGACHSVKSL 112
QY 188 OORQIOMQMTED-----ICRQVLLIGSLQGYGAPASBELDETGAASSKPLPLF 240
Db 113 -OROKGASPSREKPRPGVSYIYWRLLIGS-----GSGGASVS--LRLT 153
QY 241 SPIKPYQ---TSKTLAASSSSSSSSSGAETPKQAFPHLYHPLGSOHPRPSAGVGRSHKP 296
Db 154 SPLRPEGVRLREKTLTE-----NAL-----LGNQPR 181
QY 297 TPAPSA---LPESTQDLASPHLAFSTTGLLAAQCLGARGLENTASPG----- 344
Db 192 PEROKPCAOEVPRT-----FGSADLLEAAGRGHNRMGAVWQ 219
QY 345 ---LLKPKNGSELXYGEVWPLEKPGGRHKCRCAKVFSGDSALQIHLRSHTGERPK 400
Db 220 EPHRIILGGEPSTWDELGERLHAGEK---SFECAKCSKVYFKSSDLKHLKRTHTGERPYE 276
QY 401 CNVCGNRTTTRGNLKVHFRHNRREKYPHVQMNPRVPEHLDVYTSSGLPYGMSVPREKAE 460
Db 277 CAQCGKAFQSTSHLTOHORIH-----SGEPTY----- 303
QY 461 EEAATPGGVEKRPVASTALSTATESLTLLSTAGTATAGLRPAFNKVLAKAVEPKK 520
Db 304 ---ACP----- 306
QY 521 ADEHTPPGSEGAISGVASSTATIMQLSKIMTSLPSWALLTNHFKSTGSPFLPLCARAL 580
Db 307 -----VCGKAPRHSSSIVAHQRIHT-----AEKSFCSCEGKAF 340
QY 581 ---GASPSSETSKIQQLVEKIDROGAVAVTSAAGAPTTSAPA-----PSSSAS 625
Db 341 SHGSMISQHRKIH-----AGGRPYACAQCGRRRCRNSHLIONHERHTNG 383
QY 626 SGPNQVCICLRYLSCPRALRYHGQNGGERPEFKKVGCRASTRGNLANRFGNKAAPRA 685
Db 384 EKRPYCAICGAAPSGGSSSLFKHORVHTGERPFACQCGRAFSHSSNLQHOHLHTGERPF 443
QY 686 RAONSCPTICOKKFTNAVTLQOHVHMLGQIIPNGTALPEGGAQOE-----NG 734
Db 444 R---CVDGKAFAGAVALLSHRRIHTGER-----PFVCTQCGRAFRERPALFHHQRIHTG 495

QY 735 SEQSTVSGAGSFPQOQSOQPSPEELSEEEEE 766
Db 496 EKTVRRSRASLHPQARSVAGASSEGAPAKETE 527

RESULT 5

US-09-063-035-2
; Sequence 2, Application US/09063035
; Patent No. 6160091
; GENERAL INFORMATION:
; APPLICANT: PEUKERT, Karen; HAENEL, Frank; and EILERS,
; APPLICANT: Martin
; TITLE OF INVENTION: Myc-binding zinc finger proteins,
; TITLE OF INVENTION: their preparation and their use
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage

COMPUTER: IBM AT-compatible, 80486 processor
OPERATING SYSTEM: MS-DOS version 6.1
SOFTWARE: WordPerfect version 8.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/063,035

FILING DATE: 21-Apr-1998
CLASSIFICATION: 514

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 803 amino acids
TYPE: amino acid

TOPOLOGY: linear
US-09-063-035-2

Query Match 4.6%; Score 242; DB 4; Length 803;
Best Local Similarity 18.4%; Pred. No. 3.7e-08;

Matches 170; Conservative 83; Mismatches 305; Indels 364; Gaps 29;

QY 77 SEPRPGHNNPOYMDTH-----SNPPDSSGVPTDPTGPER--KEEES 120
Db 115 AEPATSPGNAEALATGCGDKRAKEKYATSTLSKLEQGRSTPT---GPSHDLKEERG 170
QY 121 GHPVATGTAGGGGLILASPKLGATPLRPESTRPAPPPPP---PPPPGVSGHLM 176
Db 171 GOAQAASGAEE-----QTEKADAPREPPVELKPDPTSGMAAAEAE 211
QY 177 IPLI-----LEELRVLOOQROIHOMQMTFOIC--ROYLLIGSLQGYGAPASBELPG 226
Db 212 AALSSSESEOMEVEPPAKGEEQKEQEEQEEGAPRAVKEGSOLENGEADPEENEEES 271
QY 227 TGTASTKPLPLPLSPIKPYQTSKTLASSSSSS--SSSSSGAETPKQAF---PHLYHPLGSO 282
Db 272 AGTIDS-----GQELGSEANGILRSCTGDRRESKAYGVYHKCECCGKE 314
QY 283 HFFSAGVGRSHKFTAPSPALPESTQDLASPHLAFSTTGLLAAQCLGAARGLEATAS 342
Db 315 FTHT-GNFKRHRIHNGEKFSCRECSKAFSDP-----AACKAHKHTS 357
QY 343 PGLLKPKNNGSELXYGEVWPLEKPGGRHKCRCAKVFSGDSALQIHLRSHTGERPKCN 402
Db 358 P--LKP-----YG-----CEEGKSTYRLISLNLKRRHSGEARYRCE 393
QY 403 VCGNRTTTRGNLKVHFRHNRREKYPHVQMNPRVPEHLDVYTSSGLPYGMSVPREKAE 462
Db 394 DCGKLFSTTSGNLKRHQLVNSGEKPY-----QCDYDGRS-----FSDPTSK----- 433
QY 463 AATPGGVEKRPVASTALSTATESLTLLSTAGTATAGLRPAFNKVLAKAVEPKKAD 522

Db 434 -----MRLETHDTK 444
QY 523 ENTPGSEGAISGVAESSTATLTMOLSKLMTSLPSWALLTNHFKSTGSPPL----- 573
Db 445 EHKCPHCD-----KRFNOGNLKAHLKIHADG 472
QY 574 PLCARLALGASPSFETSKLOOLVEKIDROGAVAVTSAASGAPTSAPAPSSSASSGPNOCVI 633
Db 473 PLKCRRCGKQFTTSGNLKRLR-----IHSGEKPYCIH 506
QY 634 CLRVLSCPRALRLHYGQHGGERPKCVCGRAFSTRGNLRAHEVGH----- 679
Db 507 CQRQFADPGLQHRVIRHTGKPCQCVWCCKAFQASLIANHRQHTGKERYCERCKR 566
QY 680 --KASPAAR-----AONSCPTCQKFTNAVTLQOHVYRMHLGGQIPNGCTALPEGGGA 729
Db 567 FVQSSOLANHIRHNDNIRPHKCSVSKAFVNVGDLSKHITIIHNG-----EKPYL 615
QY 730 AONGSEOSTVSGAGSFPOQSOQPSPEELSEEEDEDEEDYTDDESLAGRSESG 789
Db 616 CDKCGKGFNVNDLRSHVKTYHQGAKIKIL-----EPGSEVSVYTVDDMYT----- 664
QY 790 GEKAISVRGDSSEASGAEEVGTVAATAAGKEMDSNEKTTQOSSLP PPPPSLDQPOP 849
Db 665 -----LATEALATA-----VTQLTVVP----- 682
QY 850 MEGSSGVLGKEGKGERSSSPASALTPEGEATSVTLVEELSLDPAKKEPSESSRK 909
Db 683 -----VGAAYT--ADETEVLKAEISKAVQOVEDPNTILY 717
QY 910 ACEVCGQAFPSQALAEHOKTH 931
Db 718 ACDSGCKFLDANSLAQHVRH 739
RESULT 6
US-08-475-844-5
Sequence 5, Application US/08475844
Patent No. 5972643
GENERAL INFORMATION:
APPLICANT: Lobanenko, Victor V.
APPLICANT: Neiman, Paul E.
APPLICANT: Klenova, Elena M.
APPLICANT: Goodwin, Graham H.
APPLICANT: Philippova, Galina N.
APPLICANT: Collins, Steven J.
TITLE OF INVENTION: CTCF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourile and Crew
STREET: one Market Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,844
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/261,680
FILING DATE: 17-JUN-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14538A-11-1
TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 728 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-475-844-5
Query Match 4.6%, Score 241; DB 2; Length 728;
Best local similarity 20.8%; Pred. No. 3.8e-08;
Matches 112; Conservative 60; Mismatches 235; Indels 132; Gaps 16;
QY 365 EKPGGRKRCRCAYGSDSALQIHLRSHTGERPYKCNOCNFTTGRNLKVHFRHREK 424
Db 291 ERP--HKCHLGRGFRVTLRLNHLNHTGTRPHKCPDDMAFVYSGLY----RR-R 342
QY 425 YPHQMPNHPREHIDYVITSSGLPYGMSVPEKAEAEATPGGVERKPLVASTTALSA 484
Db 343 YKHTHEKPKC-SMCDYA-----SVEYSKLKRHISHTG--ERPGCSLCSYAS 388
QY 485 TESLTL--LSTAGTATAPGLPAENKFLVLMKAVEKRNKADENTPPGSEGAISGVAESS 541
Db 389 RDTYKLRHMRTHSGEKPEYCYICHARF-----TQSG 420
QY 542 TATLMQSLKMTSLPSWALLTNHFKSTGSPPLPCARALGASPSFETSKLOOLVEKIDROG 601
Db 421 TKNHILQK-----HTENVAKFHCHCDTVIARKSLDGLVNLKROHSTIEGK 467
QY 602 AVAATSAASGAPTSAPAPSSSASGPNOCVICLRVLSCPRALRLHYGQHGGERPKCV 661
Db 468 KCRYCDAYFHERYALIQHKSXKNEKRFKQDCDDYACROEHNVMHRTHTGKEKYACSH 527
QY 662 CGRAFSTRGNLRAHFVGHKASPAARQONSCPTCQKFTNAVTLQOHVYRMHLGGQIPNGCT 721
Db 528 CDKTFROKQLDMHFRKYDHPFVPAFVCSKCGKFTTRNTMARHA-----DNCS 578
QY 722 ALPEGGAOENGSEOSTVSGAGSFPOQSOQPSPEELSEEEDEDEEDYTDDESL 781
Db 579 GLDGGEG--ENGGETKKKRGKRRKRSKKESSDSDEANEPDLDDNDEE----- 627
QY 782 AGRGSESGEKAISVRGDSSEASGAEEVGTVAATAAGKEMDSNEKTTQOSSLP PPPPP 841
Db 628 -----ETAVEIE-----AEPEVSAEAPAPPSK-----KRGGRPKGA 661
QY 842 DSLDQPOP-----MEGSSGVLG-----KEGKPERSSSPASALTPE 880
Db 662 TOTKOSPAALIQVEDONNGEINIIYVYKKEPDAETVEEEDQAPAVEAVANGDLTPE 720
RESULT 7
PCT-US95-08429-5
Sequence 5, Application PC/TUS9508429
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CTCF
NUMBER OF SEQUENCES: 21
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08429
FILING DATE: 15-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/261,680
FILING DATE: 17-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.

REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14538A-11-1PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 728 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-08429-5

Query Match 4.6%; Score 241; DB 5; Length 728;
Best Local Similarity 20.8%; Pred. No. 3.8e-08;
Matches 112; Conservative 60; Mismatches 235; Indels 132; Gaps 16;

QY 365 EKGGRKRCPCAKVFGSDSLQHLRSHTERPYKCNVGNFRTGRLKVEHRRERK 424
DB 291 ERP--HKHLGGRAFRTVTLRLMLNTHGTRPHKCPDCDMAFVTSGLV----RRR-R 342
QY 425 YPHVQMPHPVPEHLDYVITSSGLPYGMSVPEPEKAEAEATPGGVERKPLVASTALSA 484
DB 343 YKTHHEKPRFC-SMCDYA-----SVEYSKLKRLHRSHTG---ERPQCSLCSTYAS 388
QY 485 TESLT--LSTAGTATAPGLPAFNKFLMKAVBPKNKADENTPPGSESAISGVAESS 541
DB 389 RDIYKLRHMRTHSGEPRYECYICHRF-----TQSG 420
QY 542 TATLMOLSKMTSLPSPNALTNHFKSTSSPLPLCARALGASPSSTSKLOQLEYKIDRQG 601
DB 421 TKNMHLLOK-----HTENVAKFHCPHCDYVIAKRSDLGVHILRKQHSYIDQK 467
QY 602 AAVAVTSASGAPTTSAPAPSSASSGPNQVCYICLRYLSCPRALRLHYGGHGERPEKCV 661
DB 468 KCRYCQAVFHERALALIOHSHKNEKRRKQOCDACRGERHMMVMKRTHTGEKPYACSH 527
QY 662 CGRAFSTRGNLRHFNHGKASPAARAONSCPTQCKKFTNAVTLQOHVHMLGGQIPNGGT 721
DB 528 CKTFKOKOLLDHNFKRYHPNFVPAAFVCSKCKGTFTRRNMARHA-----DNCS 578
QY 722 ALPEGGAQENSGEOSTVSGAGSPFOQSQSPPEELSEEEDEDEEEDVTDSDSL 781
DB 579 GLDGGG---ENGETKKGKGRKRRKMSKEDSDSENAEPDLDDNDEE----- 627
QY 781 AGRGSESGEKAISVRDSEASGAEEVGTVAATAAGKENDSNEKTTQOSSLPPPPP 841
DB 628 -----ETAVEIE-----AEPEVSAEAPAPPSK-----KRRGRPGKAA 661
QY 842 DSLDQPOP-----MEQGSSGVLG-----KEGGKPRSSSSPASALTPE 880
DB 662 TQTQKOPAPAIIVQEDONTGEIENIIVEVKKPEPAETVEEEEAQPAVAVEAPNGDLTPE 720

RESULT 8
US-08-459-568-2

Sequence 2, Application US/08459568
Patent No. 5811304
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,568
FILING DATE: 02-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-459-568-2

Query Match 4.5%; Score 238.5; DB 2; Length 1706;
Best Local Similarity 20.7%; Pred. No. 1.7e-07;
Matches 209; Conservative 97; Mismatches 319; Indels 387; Gaps 45;

QY 4 ESERSRLGVP-----AGEPAELGDASEEDHPQVCARKCAQFTDPTFEIAHONAC 54
DB 828 KSEGTGTPVPWESVLDLSVYKHKPCDSEKKEKENH---LAQPAKKKKRKTTCMLQK--- 881
QY 55 STDPVAVVITGGQENPNNSASSSEPRRGGHNNPQVMDTEHSN-----PDGSSVPTDPT 109
DB 882 -----VIL--NEYGVSLPTETTPETKSPCKSPDTPDPLGPDSSCSVPTAES 931
QY 110 WGPERRGEESGHLVATGTAAGGGGLILASPKIATPLPESTPAPPPPP----- 164
DB 932 -PPEVVGPSPP-----LQTAISLSSQLEPPLLTPTESPPPCPVL 972
QY 165 -----PPPGVSGHLNPLILEELRVLQROIHQOMTEQICROVLLLSLGTGVPAS 220
DB 973 TVATPPPLPLTPVPLSHP-----SSDAS 995
QY 221 PSELPP---GTGTASTPPL-----PLFSPKPYOTSKTLAS-----SSSSSSSSSGAE 266
DB 996 PQCPSPFSNTTAQSPILSPITVSPSPPLPPVPLMSAASBPPTLSSSSSSSS--- 1052
QY 267 TPKQAFPHLYHPLGSHQHFPSAGVGRSHKPTPAPSPALP---GSTDOLIAS-PLAFPT 322
DB 1053 -----FPSSSCS---STSPSPPLSAVSSVSGDNLEASLPVATFKQ- 1092
QY 323 TGLLAOCIGAAGLEATASGGLKPKNGSGELSTGEVWGPLKRP---GGRHK----- 372
DB 1093 -----EESSEEG-LKPKR-----EAPRAGQGVQETFSKN 1122
QY 373 --CRFCAKVFGSDSALOILHRSHTGERPYKCNVGNFRTGRLKLVH--PHR----- 421
DB 1123 FICNVCESPFLSIKDLTKHLSVNAEBMFKEFCVQLKVKITDLSHRLLHGVGINFVC 1182
QY 422 ---REKYPHV-QMNPHPVPEHLDYVIT-----SSGLPYGMSVPEKA 459
DB 1183 SVCKKEFAFLCNQOHRDLHPDEVCTHHEFESGLTRQNFPTDPKANEHNPSPLEPFL 1242
QY 460 E-----EAPATP---GGGV-----RKPLVAS 478
DB 1243 ETSREELNDSSSELYTTIKIMASGIRTKDPDVRGLGNQHHYPSFKPPFPQYHHRNPMGIG 1302
QY 479 TVALS-ATESLTLSTAGATAPG-----LPANFKFVL-----MAVEKKNK---AD 522
DB 1303 VTATNFTTNINPTFTTALICTGCGKGVDMNPPELHKHLACASASDKKRYTTRKKNPVPIK 1362
QY 523 ENTTPGSESAISGVAESSATLMQLSKMTSLPSPWALLTNHFKSTGSPPLPLCARALGA 582

Db 1363 QTVQPRNGVAVLDNSGKNAFRRMGOKPRLSEFVNLGKMSPNKTL----- 1407
QY 583 SPSETSKLQOLVEKIDROGAVAVTSAAGAPTTSAAPSSASSGPNOCVLCR----- 636
Db 1408 --SALKKKNLQVQK-----AIIQKNRAKOKADRLDTSEASS--HICPYCDREFTYIG 1456
QY 637 -----VLSGP-----RALRLHYGQHGGER----- 655
Db 1457 SLNKHAAFCSPKPKPLSPSKRKRVSHSSKKGGHASSSSSDRNSSCHPRRRRTADTEIKMSTQ 1516
QY 656 -----PFKCYGCGRAFTRGNLR-AHFVGHKASPARAONSCPICQKFTN-- 700
Db 1517 APLGKTRARSTGPAQASLPSSSFRSQNVKFAASVSKKASSSLRNSSPIRMAKITHVE 1576
QY 701 ----AVTLQO-----HYRMHLGQIIPNGGTALEGGGAQENGSEOSTV--- 740
Db 1577 GKPKRAVAKSHSAQLSSKSGRLHVQKSKAVIOSKTAL-----ASKRRTRDFIVKSR 1630
QY 741 --SGAGSFPQOQSOPSPPEELSEEEDEDEDEE--DYTDEDSLARGSES 788
Db 1631 ERSGG--PIRSLQLAAADLSESRREDSARHLEKDFSYSLRLASRCSS 1679

RESULT 9
US-08-399-411-2
Sequence 2, Application US/08399411
Patent No. 5831008

GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/399,411
FILING DATE: 06-MAR-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-399-411-2

Query Match 4.5%; Score 238.5; DB 2; Length 1706;
Best Local Similarity 20.7%; Pred. No. 1.7e-07;
Matches 209; Conservative 97; Mismatches 319; Indels 387; Gaps 45;

QY 4 ESERSSRLLGVP-----AGBPALGSDASEEDHPQVCAKCAQFTDPFEFLAHQNAK 54
Db 828 KSEGKGRKPPVPEASVLDLSVHKKPCDSEGEKFEKENH---LAQPAKAKKKPTTCMLQK--- 881
QY 55 STDPVAVIIGQENPNSSASSEPRPEGHNNPQVMDTEHSNP-----PDGSSSVPTDPT 109

Db 882 -----VLL---NEYNGVSLPTETTPETVTRSPSPCKSDPTQDPDELGPDSSCSVPTAES 931
QY 110 WGPERRREESGHLLVAATGTAGAGGGGLILASPKLGATPLPESSTAPRPPPP----- 164
Db 932 -PPEVVPSSPP-----LQTAASSGQLPPLPTPETEPPSSPPCPVYL 972
QY 165 -----PPPPGVSGHLNPLILEELRVLDQRIHOMQTEQICQVLLIGSLGQTVGAPAS 220
Db 973 TVATPPPLPTVPLSHP-----SSDAS 995
QY 221 PSELPP---GTGTASTKPLL-----PLFSPIKPVQTSKTLAS-----SSSSSSSSSGAE 266
Db 996 PQQCPSPFSNTTAQBPILPILSPVSPSPPIPVPEPLMSAASGPPPLSSSSSSSSS--- 1052
QY 267 TPKAQFHLIPLPGSQHPSPSAGVGSRSHKPTAPSPALP---GSTDLLAS-PLHAFEST 322
Db 1053 -----FPSSSGS---STSPSPPLSAVSSVSGDNLASLPVTFKO- 1092
QY 323 TGLLAQCLGAARGLAATASPGLLKPKNGSGELSYGEVMGPLKRP--GGRHK----- 372
Db 1093 -----EESSEGLKPKK-----EAPRAGGQSVVQETFSKN 1122
QY 373 --CRFCARVFGSDSALQILRSHTGERPYKCNVCGNRFTTGNLKVH-FNRH----- 421
Db 1123 FICNVCESPFLISKDLTKHLVHAHEMPKCEFCVOLFKKYKTDLSEHNFLLHGVGNIFVC 1182
QY 422 ---REKYPHV-OMNHPVPEHLDVYT-----SSGLPYGMSVPRKA 459
Db 1183 SVCKKEFAFLNLOOHORDLHDEVCTHHEFESGTLRPNQFTDPSKANVEHMPSLPEBL 1242
QY 460 E-----EEAATP---GGGVE-----RKPLVAS 478
Db 1243 ETSREELNDSSEELTYTIKIMASGITKDPDVRGLNGHYPSPKRPFOYHNRHPMGIG 1302
QY 479 TTAALS-ATESLTLLSTAGTAPG-----LPAPNKFVL-----MKAVEPKK---AD 522
Db 1303 VTATNFTTHNIPQFTTTAIRCTKCGKGVDMPELHKHILACASASDKKRYTPKKNPVPLK 1362
QY 523 ENTSPGSEGSALISGVAESSTATLMOLSKLMTSLPSMALTRHNFSTGSEFPLPCARALGA 582
Db 1363 QTVQPRNGVAVLDNSGKNAFRRMGOKPRLSEFVNLGKMSPNKTL----- 1407
QY 583 SPSETSKLQOLVEKIDROGAVAVTSAAGAPTTSAAPSSASSGPNOCVLCR----- 636
Db 1408 --SALKKKNLQVQK-----AIIQKNRAKOKADRLDTSEASS--HICPYCDREFTYIG 1456
QY 637 -----VLSGP-----RALRLHYGQHGGER----- 655
Db 1457 SLNKHAAFCSPKPKPLSPSKRKRVSHSSKKGGHASSSSSDRNSSCHPRRRRTADTEIKMSTQ 1516
QY 656 -----PFKCYGCGRAFTRGNLR-AHFVGHKASPARAONSCPICQKFTN-- 700
Db 1517 APLGKTRARSTGPAQASLPSSSFRSQNVKFAASVSKKASSSLRNSSPIRMAKITHVE 1576
QY 701 ----AVTLQO-----HYRMHLGQIIPNGGTALEGGGAQENGSEOSTV--- 740
Db 1577 GKPKRAVAKSHSAQLSSKSGRLHVQKSKAVIOSKTAL-----ASKRRTRDFIVKSR 1630
QY 741 --SGAGSFPQOQSOPSPPEELSEEEDEDEDEE--DYTDEDSLARGSES 788
Db 1631 ERSGG--PIRSLQLAAADLSESRREDSARHLEKDFSYSLRLASRCSS 1679

RESULT 10
US-08-516-859A-2
Sequence 2, Application US/08516859A
Patent No. 6069231
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 106

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: Campbell & Flores LLP
3 STREET: 4370 La Jolla Village Drive, Suite 700
4 CITY: San Diego
5 STATE: California
6 COUNTRY: USA
7 ZIP: 92122
8
9 COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: Patentin Release #1.0, Version #1.25
14
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/516,859A
17 FILING DATE: 18-AUG-1995
18 CLASSIFICATION: 514
19
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: US 08/399,411
22 FILING DATE: 06-MAR-1995
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: US 08/292,683
25 FILING DATE: 18-AUG-1994
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Campbell, Cathryn A.
28 REGISTRATION NUMBER: 31,815
29 REFERENCE/DOCKET NUMBER: P-LJ 1776
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: (619) 535-9001
32 TELEFAX: (619) 535-8949
33 INFORMATION FOR SEQ ID NO: 2:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 1706 amino acids
36 TYPE: amino acid
37 TOPOLOGY: linear
38
39 MOLECULE TYPE: protein
40 US-08-516-859A-2

Query Match 4.5%; Score 238.5; DB 3; Length 1706;
Best Local Similarity 20.7%; Pred. No. 1.7e-07;
Matches 209; Conservative 97; Mismatches 319; Indels 387; Gaps 45;

4 ESESSRLGVP-----AGEPAELGDSSEHDPVCAACCAQFTPTTELAHONAC 54
828 KEEGKGTTPWESVLDSVHKKPCDESGKEFENH---LAOPAAKKKKPTTCLMOK--- 881
55 STDPVAVLIGCOENPNSSASSEPRRGHNNPQVMDTEHSP-----PDSGSSVPTDPT 109
882 -----VLL--NEYNGVSLPTTETTEVTSPSPCKSPDTPQDELGPDSSCSVPTAES 931
110 WQPERGESSGFLVAATGTAAAGGGGLIASPKLGATPLPESTPAPPPPP----- 164
932 -PPEVVGPSPP-----LQTASLSSGQLPRLPTPESSPPCPVVL 972
165 -----PPPGVSGHNLPLLELRVLQQRQIHQMOMTEQICROYULLGSLQTVAPAS 220
973 TVATPTPLPLTPVPLSHP-----SSDAS 995
221 PSELP--GTGTAASVTKPL-----PLFSPIKPYQTSKTLAS-----SSSSSSSSGAE 266
996 PQQCSPPFNTAOSPLPIPSPIVSPSPPIPIVPLMSAASPGPTLSSSSSSSSSS--- 1052
267 TPQKAFHLIYHPLGSHPPSAGVGGRSHKTPPADSPALP---GSTDQLIAS-PIHAPST 322
1053 -----FPSSSCS---STSPSPPLSAVSVSSGDNLASLPAVTFKQ- 1092
323 TGLLAOCIGAARGLEPAASPLGLPKNGSGELSYGEVNGLEKP--GGRHK----- 372
1093 -----ESESSEG-LKPK-----EAPPAQGGQSVVQETFSKN 1122
373 --CRCFAVFGSDSALOIHLSRHTGERPKYKCNVCGNRTTRGNLKVH-FHRH----- 421
1123 FICVNCESPFSLIKDLTKHLVSAAEEMPFCEFCVQLFKVATDLSSEHRLHLHGVCNIFVC 1182

QY 422 ---REKYPHV-QMNPYPPEHIDYIT-----SSGLPYGMSVPEPKA 459
DB 1183 SVCKKEPAFLNLOQHODLHDEVCSTHEFESGLRPQNFTDPSKANVEHMPLEPEPL 1242
QY 460 E-----EEAATP-----GGGVE-----RKPLVAS 478
DB 1243 ETSREELNDSSEFLYTIKIMASGIKTKDPPVRLGLNQHYPSPKPPFPQYHHNPNMGIG 1302
QY 479 TTAIS-ATESLTLTSTAGTATAPG-----LPAENKVL-----MKAVEPKNK---AD 522
DB 1303 VTAVNFTTHNIPQFTTTAIRCCKCGKGVDMPELHKHILACASASDKRYTPKKNPVPLK 1362
QY 523 ENPPGSEGSAISGVAESSTATIMQSLKMTSLPSALLTNHKKSGSPFLPLCARALGA 582
DB 1363 QTVQPKNGVVVLNDSGKNAFRMRGQPKRISFVNELGKMSPNKIKL----- 1407
QY 583 SPSETSKLOQLVERKIDROGAVAVTSAAGAPPTSAAPAPSSASSGPNOCYICLR----- 636
DB 1408 --SALKKKKNQLVOK-----AIIQKNRAKOKADLRDTSEASS--HICPYCDREFTYIG 1456
QY 637 -----VLSCP-----RALRLHYGHGGER----- 655
DB 1457 SLNKHAAFSQPKKPLSPSKRKRVSHSKKGGHASSSSSDRNSSCHPRRTADTEIKMOSTQ 1516
QY 656 -----PPKCKVCGRAFTRGNLR-AHFVGHKASPARAONSCPIQKKFTN-- 700
DB 1517 APLGKTARSTGRPAOASLPSSSFRKRONVKFAASVSKKRASSSLRNSSPTIRAKTTHVE 1576
QY 701 -----AVTLQO-----HYRMHGGQIPNGTALPEGGAAGENGSEOSTV--- 740
DB 1577 GKRPKAAVAKSHSAQLSSKSGRLHVRVOKSKAVIQSKTAL-----ASKRRTDRIYKSR 1630
QY 741 --SGAGSFPOQOQSPSEELSEEEDEDEEE--DVTDEDSLARGSEES 788
DB 1631 ERSGG---PITRSLQLAAADLSERREDSSARHKLDFSYSLRLASRCGSS 1679

RESULT 11
US-09-586-472-2
Sequence 2, Application US/09586472
Patent No. 632335
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
Zinc Finger Proteins
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/586,472
FILING DATE: 01-Jun-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/528,706
FILING DATE: 17-MAR-2000
APPLICATION NUMBER: US 08/516,859
FILING DATE: 18-AUG-1995
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.

Db 932 -PPEVVGSSSP-----LQTASLSSGQLPPLLPTEPSSPPCPVL 972
QY 165 -----PPPGVGSGLNPLILEELRVLOQRQHOMQMTQICQVLLGLGTVGAPAS 220
Db 973 TWAATPPPLPPLPVLSHP-----SSDAS 995
QY 221 PSELPL---GTGTASSTRPL-----PLFSPKPVQTSKLAS-----SSSSSSSSSGAE 266
Db 996 PQQCPSPFSTMTSPILSPVSPSPSPPIPVPEPLMSABGPPTLSSSSSSSSS--- 1052
QY 267 PKQAFHLLHYPLGSGQHPFSAGVGRSHKPPAPAPALP---GSTDQLIAS-PLAPPST 322
Db 1053 -----FPSSSCS---STSPSPPLSAVSSVSSGNDLXSLPAVTFKQ- 1092
QY 322 TGLLAQCLGANGLENTASPGILKPKNGSELSTGEVWGPLEKP---GGRHK----- 372
Db 1093 -----ESESSEG-LKPYE-----EAPAGQGVVQETFSKN 1122
QY 373 ---CRPQAKVFGSDSALQIHLRSHTGERPKVCNCGNRFTTGNLKVH-FPHR----- 421
Db 1123 FICNVGESPFLSKIDLTKHLSVHAEMPFKEFCYQLPKVKTDLSHRPLRHGVGNIFVC 1182
QY 422 ---REKYPHV-QMNPHPVPHLDVYT-----SSGLPYGMSVPEKA 459
Db 1183 SVCKKKEFAFLCNDLQHQRDLDHDEVCTHHEFESGTLRPQNFDPDSKANVEHMSLPEPL 1242
QY 460 E-----BEAATP-----GGGVE-----RKPLVAS 478
Db 1243 ETSREELNDSSSELYTTIKIMASGKITKDPDVLGLNQHYPSPKPPFOYHNHNPWGIG 1302
QY 479 TTALS-ETSLTLLSTAGTATAPG-----LPAFKFVL-----MKAVERPKN---AD 522
Db 1303 VYATNLTTHNIPTFTTAICTGCGKGVNDMDELKHLILACASADSKKRYTPKKNVPLK 1362
QY 523 ENTPEPDESAISGVASSATITLMQLSKMTSLPSWALLTTHFKSTGSPPLICARALGA 582
Db 1363 QTVQPPGAVVLDNSGSKNARMQOPKRLSENVELGKMSPNLKL----- 1407
QY 583 SPSESKTLQIVLEKIRQGAVAVTSAASGAPTTAPAPSSSSSGPNOCVICLR----- 636
Db 1408 ---SALKKKNOVLQK-----ALIQKRNAAKQKADLRDTSEASS--HICPYCDREFTYIG 1456
QY 637 -----VLSCP-----RALRLHYGQHGGER----- 655
Db 1457 SLNKHAAFECPPKPLSPSKRKVSHSSKKKGHASSSSSDRNSCHPRRTADPEIKMQSTQ 1516
QY 656 -----PFKKVCGRAFSTRGNLR-AHFVGHKASPARAQNOSCPIQCKKFTN-- 700
Db 1517 APLKTRASTGPAQASLPSSSFRSRONVAFKFAASVSKKASSSLRNSSPIRMAKITHVE 1576
QY 701 ---AVTLQO-----HYRMHLGQIIPNGTALPDPEGGAOENGSEOSTV--- 740
Db 1577 GKPKRAVAKSHSAQSSKSSRGHLHVQOKSKAVIQSKTAL-----ASKRRIDRFIVKSR 1630
QY 741 ---SGAGSPFOQOQSOPSEPELSEEEDEEBEER--DVTDEDSLGRGSES 788
Db 1631 ERSQG---PIRSLQLAAADLSESRREDSARHELDKFSYSLRLASRCSS 1679

RESULT 13

JS-08-246-489-2

Sequence 2, Application US/08246489

Patent No. 6225049

GENERAL INFORMATION:

APPLICANT: Ian, Michael S.

APPLICANT: No. 6225049kins, Abner L.

TITLE OF INVENTION: NOVEL HUMAN INSULINOMA-ASSOCIATED CDNA

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobb, Martens, Olson & Bear

STREET: 620 Newport Center Drive

CITY: Newport Beach

STATE: California
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/246,489
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/901,715
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH012.012A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 510 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-246-489-2

Query Match 4.3%; Score 225.5; DB 4; Length 510;
Best Local Similarity 23.3%; Pred. No. 2.6e-07;
Matches 127; Conservative 46; Mismatches 232; Indels 135; Gaps 23;

QY 133 GG--GGGLIASPKLG---ATPLPEESTPAPPPPPP-----PPPP 168
Db 22 GGGDGRALLSPSCGGAAREPAPSPVPGPLPPPPAPRAHAAALAAALACAPGPDPPO 81
QY 169 GVSGSHLN-----IPLILEELRVLOQRQHOM-----QMTQICQVLLGSL 211
Db 82 GPRAAHFGNPEAHPALVLPPTRVSRHEHKKYERSEFNLGSPVSAESFPPTALLGGG 141
QY 212 GQTVGAPASSELPGTASTASTKPLPLFSPKPVQTSKLASSSSSSSSSSAETPKQA 271
Db 142 G---GGGASGA---GGGTCGGDPL-LFAPAEIKMGTAFSAGAEAAKRGPGPPLPRAA 193
QY 272 FPHLYHPLGSHHPFSAGVGRSHKPTPAPSPALPGSTQDLIASPHLAFSTTGLAAQCL 331
Db 194 ---ALRPCKRRP-----PPTAAERPA-----KAVAPCAKAKK----- 224
QY 332 GAARGL-----EATASPOL-LKPKNGSELSTGEVWGPLEKPGKRCFACAVFGSDAL 386
Db 225 -AIRKLHFDEVTTSPLVIGLIKIKEGPVEAPRGRA-GGAARPGEFTIQCLCKEYADPFL 282
QY 387 QHLRSHTGERPKVCNCGNRFTTGNLKVHFNHREKLYRNQMPHPVPEHLDVYTSS 446
Db 283 AQHKCSRIYRVEYKCECAAVFSCPANLASHRRWKKRP-PAPAAARAPPE-----AA 334
QY 447 GLPYGMSVPEKAEBAATPFGGVE-----RKPLVASTTALS 483
Db 335 ABAEAREAPGGGSDRDTSPSGVSESESDGLYBCHNCAKKFRQAYLRKHLAHNQALQ 394
QY 484 ATESLTLLSTAGTATAAPGILPAFNKFLVLMKAVEPKKADENTPPGSESAISGVASSSTA 543
Db 395 A-----KGAPLAP--PAEDLLALVPG--PDEKAPQEAAGDGEAGVLGLSASABC 440
QY 544 TLMQLSKMTSLPSWALLTTHNFK-----STGSPPLICARALGASPESTKIQOLVLEKIDRQ 600
Db 441 HICPVCG---ESFRASKGAQERHRLHLHAQVFPCKYCPATPYTSSPGLTTHINKCHPSENKQ 498

```
RESULT 14
US-09-579-181-2
; Sequence 2, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chivita, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2972
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-2

Query Match          4.3%; Score 225.5; DB 4; Length 2972;
Best Local Similarity 21.0%; Pred. No. 2.5e-06;
Matches 225; Conservative 101; Mismatches 366; Indels 377; Gaps 50;

OY 58 PVMVITIGQENPNNSASSEPR-----PEGHNN---POVMDTEHSNPPDSGSS 103
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 890 PSLGLVLSGTSRPTPTLSLKPTRPAPVRLSPAPRPGSSSLKPLTVPPGYTFPPAATTT 949
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 104 VPRD-----PTWGPR-----RGE-ESSGHFLVAATGTAAGGGG-- 137
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 950 TSTTTTATTATTAVPAPTPAPORLLISPDMAHLPSGEVVSIGOLASLAORPVANAGGSKP 1009
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 138 -----LILASPKLGATPLPPES 154
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1010 LTFQIOGNKLLTGAOVROLAVGPRPLQMPPTMVNNTGVYKIVVRQAPRDLTPVP--L 1068
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 155 TPAPPP-----PPP-----PPRGVSGHLNIPILIELRLVLOQRQIHQMOMTE 199
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1069 APAPRPPSSGLPAVLNPRPTLTGRLPTTLGTARAAMP-----TP 1109
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 200 QICROVLL-----GSLGQTVGAPAS--PSELPGTGASTTKPL---LPLFSP 242
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1110 TLVRPLKLIVHSPREVSASAPGAAPLTITSSPLHVSSLP--GASSPMPIPNSSPLASP 1167
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 243 IKPVQTSKTLASSSSSSSSSSGAET--PKQAFHLVHPILGSHPEFSAGVGRSHKPTPAPS 301
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1168 V-----SSTVSVPLSSSLIPISVPTLPAAPASAPLTIPISAPLTVSASGAPALLTSVTPPLA 1222
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 302 PALPGSTDLIASPHLAPPTTGLLAOCLGARGLEATASPG--LLKPKNGSGELSTGE 359
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1223 PVYPAAGPPSLQPGASPSASALTGLG--LATAPSLSSSQTPGHPLLLAPTSNHVGLNS 1280
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 360 VMGLEKPGGHNKCFKCAKVFSGDSALQIHLRSHTGERPYKCNVCNFTTRGNLKVHFH 419
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1281 TVARA-----CSVYLVPASAL----- 1296
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 420 RHREKYPHVOMNPHRVEHLDYVITSSGLPYGMSVPEPEKAEAEATPGGVERKLVAST 479
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1297 --ASPFPSA--PNPAPAA-----SLAPASSASQALATP-----LAPMAAPQ 1335
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 480 TALATBELTILSTAGRTATAPGLPAFNKFLMKKAVEPKKNADENTPGSGSALISGV-- 537
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1336 TAILAPSPAPPLAPLVLPASPGAAP--VLASSQTPVPMADSPSTGTSLASASPVPA 1391
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 538 -----AESSTATLMOLSKIMTSLPSWALLNHFKSTGSPFL--PLCARAL--GASPESTKL 590
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1392 PTPVLAPBSTQIMLP--APVPSPLPSA-----STQTLALAPALAPTLGGSSPQOTLSL 1443
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 591 QQLVEKIDROG-----AAVNTSAAGAPPTS-----APASSASSGPNQVCYICLRVLSC 640
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1444 ----GTGNQGGPFPPTQTLTLTPASSLVPTRPAQTLSLAPGP-----PLGPTQ-----TLSL 1489
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

```
OY 641 PRALRLHYGQHGEBRPPCKYCGRAFTSRGNLRAHFVGHKASPARAONSCPICQKKFTN 700
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1490 APAPPLPASPVPAP-----ARTL-----TLAPASSASLILAPASVQ 1527
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 701 AVTLQOHVRMLHGOIPINGTALPEGGAAQENGSEOST--VSGAGSPQOQSQPSPEE 758
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1528 TULTSPAPVPTLG--PAAQTLALAPASTQSPASQASSLVVASGA-----APLPVT 1577
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 759 ELSEEEEEDEEEDVDYDESLAGR--GSESGEKALISRGDSDEASGAEETVAAAA 817
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1578 MVSRLPVSKDE-----PPTLTLRSGPPSPSTSTSGGPPRRQ----- 1616
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 818 TACKEMDSNEKTTQSSLP PPPP-----DSLDPQPMQ-----GSSGV 857
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1617 -----PPPPRSPFLTDLSEKKRKNORSRLERIFQLSEAHGALAP 1657
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 858 LGKKEE--GKKPERSSPASALTP-----EGEATSVTLVEEL--SLQEMRK- 900
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1658 VYGEVLDPFCFLPQPVASPIGPRSPGSHPTWTYEAHRAVLFPQOARDLQLEIIEFR 1717
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 901 ----EPGES--SSKACEVCGAAPSQALDEH--QKTHKKEPLFTCV 941
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1718 IFVMPVEAPPPSLHACHPPWMLAPROAAFOEOLASELMRAPRLHRIV 1766
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

```
RESULT 15
US-09-579-181-1
; Sequence 1, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chivita, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3118
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-1

Query Match          4.3%; Score 225.5; DB 4; Length 3118;
Best Local Similarity 21.0%; Pred. No. 2.6e-06;
Matches 225; Conservative 101; Mismatches 366; Indels 377; Gaps 50;

OY 58 PVMVITIGQENPNNSASSEPR-----PEGHNN---POVMDTEHSNPPDSGSS 103
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1036 PSLGLVLSGTSRPTPTLSLKPTRPAPVRLSPAPRPGSSSLKPLTVPPGYTFPPAATTT 1095
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 104 VPTD-----PTWGPR-----RGE-ESSGHFLVAATGTAAGGGG-- 137
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1096 TSTTTATTATTAVPATTPAPQRLILSPDMQARLPSGEVVSIGOLASLAORPVANAGGSKP 1155
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 138 -----LILASPKLGATPLPPES 154
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1156 LTFQIOGNKLLTGAOVROLAVGPRPLQMPPTMVNNTGVYKIVVRQAPRDLTPVP--L 1214
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 155 TPAPPP-----PPP-----PPRGVSGHLNIPILIELRLVLOQRQIHQMOMTE 199
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1215 APAPRPPSSGLPAVLNPRPTLTGRLPTTLGTARAAMP-----TP 1255
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 200 QICROVLL-----GSLGQTVGAPAS--PSELPGTGASTTKPL---LPLFSP 242
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1256 TLVRPLKLIVHSPREVSASAPGAAPLTITSSPLHVSSLP--GASSPMPIPNSSPLASP 1313
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 243 IKPVQTSKTLASSSSSSSSSSGAET--PKQAFHLVHPILGSHPEFSAGVGRSHKPTPAPS 301
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

```
Db 1314 V-----SSTVSVPLSSSLPISVPTTLPLPAPASAPLITIPISAPLITVSASGPAALTSTPPLA 1368
OY 302 PALPGSTDOIILASPHLAPSTTTGILAAOCLGARGLEATSPG--ILKPRNGSELSTGE 359
Db 1369 PVVPAAGPSPLOSPOSGASPSASALTG--LATAPSLSSSOTPGHPLLLAPTSTVHPGINS 1426
OY 360 VMGPLEKPGGRHKRCFCAKAVGSDSALOIHLSRHNGERPCKVNCGNRFTTRGNLKVHFH 419
Db 1427 TVAPA-----CSPVLVPAASAL-----1442
OY 420 RHHEKYPHOMNHPVNEHLDYVITSSGLPYCKSVPPKEAEPEAATPGGVEKKPLVAST 479
Db 1443 --ASPPSA--PMPAPAQA-----SLAPASSASQALAPQ-----LAPMAPO 1481
OY 480 TALSTATESLTLSTASGATAPGLPAFNKFLVMKAVEPKNKADENTPPGSEGSASCV-- 537
Db 1482 TALAPSPAPPLAPPLAPLAPSPGAAP-----VLASSOTPVVMAVPSSTPGTSLASASVPA 1537
OY 538 -----AESSTATLMQSLKMTSLPSSWALLTNHFKSTGSEPL-PLCARAL-GASPSSETSKL 590
Db 1538 PVPVLPASSSTQTMLP-APVSPPLPSPA-----STQTLALAPALPVLGSSPSQTLSL 1589
OY 591 QOLVERIDROG-----AVAVTSASGAPTS-----APAPSSASSGPNOCVLCRLVSC 640
Db 1590 -----GTGNPQGGPPTQTLSLTPASSLVPTPQOTLSLAPGP--PLGPTQ-----TLSTL 1635
OY 641 PRALRLHYGOGGERPPKCKVCGRAFTRGMLRAHFVGHKASPARARONSCPICOKKFTN 700
Db 1636 APAPPLAPASPVGPAP-----AHTL-----TLAPASSASLILAPASVQ 1673
OY 701 AVTLQOHVNRHLGQIPINGSTALPEGGGAOENGSEOT--VSGAGSPFOQSOQPSPEE 758
Db 1674 TLTLSAPVPTLd--PAAAOITALAPASTQSPASQASLVVASAGA-----APLPTV 1723
OY 759 ELSEEEDEDEEEDVTDSDSLAGR--GSESGEKAISVRGSESEASGAEEVGYVAAA 817
Db 1724 MVSRLPVSNDE-----PDTLTLKSGPPSPPTATISFGGRRPRKO-----1762
OY 818 TAGKEMDSNEKTTQOSSLP PPPP-----DSLDPQPMQ-----GSSGV 857
Db 1763 -----PPPRSPFYLDLSLEKKRRKRQSERLERIRQLSEAHGLAP 1803
OY 858 LGKKEE--GKPERSSSPASALTP-----EGEATSVTLVEEL--SLQEMRK- 900
Db 1804 VGTGYLDCTLPQVAVASPIGSPSPGSHPTFTWYTTTEAHRVNLFPQOQLDOLSEITIERP 1863
OY 901 ---BPGES--SSRKACEVGOAFPSQALAEEN--OKTHPKEGPLFTCV 941
Db 1864 IFVMPVPEAPPPSLHNCPPRPMWLAPRQAPOBOLASELMPRARPLHRTV 1912
```

RESULT 16
US-08-040-548-1

; Sequence 1, Application US/08040548
; Patent No. 5763209

; GENERAL INFORMATION:

; APPLICANT: Sukhame, Vikas P.

; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO THE

; NUMBER OF SEQUENCES: 67
; TITLE OF INVENTION: FUNCTIONAL DOMAINS OF DNA BINDING PROTEINS

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: 321 No. 5763209th Clark Street, Suite 800

; CITY: Chicago

; STATE: Illinois

; COUNTRY: U.S.A.

; ZIP: 60610

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

```
APPLICATION NUMBER: US/08/040,548
;
; FILING DATE: 514
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coughlin, Daniel F.
; REGISTRATION NUMBER: 36,111
; REFERENCE/DOCKET NUMBER: arc0067
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 744-0090
; TELEFAX: (312) 245-4961
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-040-548-1
```

Query Match 4.3%; Score 224.5; DB 1; Length 533;
Best Local Similarity 23.08; Pred. No. 3.2e-07;
Matches 135; Conservative 60; Mismatches 190; Indels 201; Gaps 28;

```
OY 99 DSGSVPTDPTWGERGEE-----SSG--HFLVAA-----TGTAAAGGGGLILA 141
Db 18 DPGSPFPHSPMTDMYPKLEEMMLLSNGAPQFLGAAGTPBGSNGSSSTSSGGGG--G 74
OY 142 SPKLGATPLPEESTPAPAPPPPPPPPPPPGVSGHMLPILLELRVLOQROIHOMQMTQI 201
Db 75 GSNSGSAFNPQGBSPD-----YEHLTTE--SESDILNNEKAMVETSYSPSQ 122
OY 202 CROVLLIGSGQVYGAAPASPELPGTGTASSTKPLLPFSPIKPY--QTSKTLASSSS 258
Db 123 TR--LPRTTYGREFLEPA--ENSGTLMPEPLPSLVSGLYSMNPTSSSASPSAA 176
OY 259 SSSSSGAETPKQAEFHLNHPILGSHQPFSGAGVGRSHKPT-PAP-----SPALPGST 308
Db 177 SSSSSASQSP-----PLSCAVPSNDSPITYSAAPFPFPNTDIFEPDQQAFAFGSA 227
OY 309 DQILASPHLAFPTTG-----LLAAQ-----CLGAA-----RGLF-ATASPL-- 345
Db 228 GTALQYPPPAVPATKGGFOVPMIDYLFPPQOGDLSLGTDPQKPPQGLNRRNQPSLTP 287
OY 346 ----LKPNGSGEL-----SYGEVM-----GPLEKPGG-----369
Db 288 STIKAFATQSGQDKALNTTYQSLIKPSRMKKYRNPSKTRPHRRPYACPYESCDRRF 347
OY 370 -----RH-----KRCFCAKAVGSDSALOIHLSRHTGERPKYKVCNGNRETTTG 412
Db 348 SRSDLLTRHTRHTGOKPFOCRICMRNFSRSDHLTTHTRHTGKERPFACDICGRKPARSD 407
OY 413 NLKVHFHRH-REKYRPHQNMNHPVNEHLDYVITSSGLPYCKSVPPKEAEPEAATPGGVE 471
Db 408 EKKRHHTKIHROK-----DKKAP-----425
OY 472 RPELVASTALSTATESLTLSTASGATAPGLPAFNKFLVMKAVEPKNKADENTPPGSE- 530
Db 426 -KSVASPAASS-----LSYSPVATSYSPATTSF-----PSVPTYSYSPGSSST 471
OY 531 --GSAISGVAESSTALIM-----QLSKLMTSLPSWALLTNHFKSG 569
Db 472 YPSPAHSGFPSPSVATTFASVPPAFTPOVSSFPAGVSSSSTSTSG 517
```

RESULT 17

US-08-466-344-1

; Sequence 1, Application US/08466344

; Patent No. 5773583

; GENERAL INFORMATION:

; APPLICANT: Sukhame, Vikas P.

; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO THE

; NUMBER OF SEQUENCES: 67
; TITLE OF INVENTION: FUNCTIONAL DOMAINS OF DNA BINDING PROTEINS


```

GENERAL INFORMATION:
APPLICANT: Mercola, Dan
APPLICANT: Adamson, Eileen D.
TITLE OF INVENTION: Inhibition of the Mitogenic Activity of
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/224,482
FILING DATE: 07-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-ME 9913
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-224-482-2

Query Match 4.2%; Score 222.5; DB 2; Length 496;
Best Local Similarity 22.8%; Pred. No. 4e-07;
Matches 127; Conservative 58; Mismatches 180; Indels 191; Gaps 26;

112 PERGEESGHFLVATGTAAAGGGGLIASPKIGATPLPESTPAPPPPPPPPPG 171
18 PEGSGNGS-----SSSSGGGG--GSGNSGSAFNFQGESEOP----- 56
172 SGHLNPLLEELRLVQROIHQMOMTEQICROYLLGSLGQTVGAPASPSLPGTAS 231
57 YEHLTTE-SFDIALNNEKAMVETSPSQTR---LPPIYTGKFSLEPA--PNSGNL 109
232 STKRLPLPESPIKPV---QTSKTLASSSSSSSSGAEPPKQAFPHLYHPLGSQHPFSAG 288
110 WPEPLFSLVSGVSMTPPTSSSSAPSPASSSSSSASQSP-----PLSCAVPSNDS 160
289 GVGNSHKPT-PAP-----SPALPGSTLDLAPSLAPPSPTG-----LLAA 328
161 SPISAAFTFPPTNDIFEPGQAFPGSAGTALQYPPPAIPATKGGQVPPMIPDLPQ 220
329 Q-----CLGAA-----RGLE-ATVAPGL-----LKPKNSSGEL-----SYGEVM----- 361
221 QQGDLISCTPDQKPGCKLENRTQPSILPLSTIKAFAPNGSQSODLKALNTYQSLIKPS 280
362 -----GPLEKPG-----RH-----KCRFCAKVGCS 382
281 RMKRYPNRPSSTPPHERPACPVESCDRRFSRDELTIRHIRIHGQKPKFOCRICKMNSR 340
383 DSALQIHLSHTGEPRIKYNVCNRPFTTNGNLKVHFHRH-REKYPHVQMNPHVPEHLDY 441
341 SDHLTTHRTHTGEPKACDICKGRKARSDERKRHTYIHLRK----- 383
442 VITSSGLPYGMSVPEKAEAEATPGGVERKPLVASTALSTATESLTLLSTACTATAP 501
384 -----DKRAD-----KSVASPAASS-----LSSYSPVATISY 411
```

```

QY 502 GLPAFNKFLMKAVEPKKADENTPGSF---GSAISVAESSTATLM-----QLSKLMT 553
DB 412 PSPATYSF-----PSPVPPSYSSPGSSTYPSAHSGPPSPVATTFASVPPAPPTQVS 464
QY 554 SLPSWALLTNHFKSTG 569
DB 465 SPFSAGVSSSFSTSTG 480

RESULT 20
US-08-459-568-4
; Sequence 4, Application US/08459568
; Patent No. 5811304
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,568
FILING DATE: 02-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-459-568-4

Query Match 4.2%; Score 222; DB 2; Length 1719;
Best Local Similarity 17.8%; Pred. No. 2.1e-06;
Matches 232; Conservative 142; Mismatches 425; Indels 502; Gaps 53;

9 SRLGVPAPEPAELG-----GDASEED-----HPVCAKCAQGTDPTE--F 47
40 TRIGVWATPILKKGKFGFVGGKSKRSQYKNNVYMEVYIPNL-GWMCIDATDEBKGW 98
QY 48 LAHQN-ACS-----TDDPV-----MTIIGQENPNSSASSEPRPE 82
DB 99 LRYVMAGSGEONLFLPLEINRAIYYKTLKPIAGEELLVYNGEDNDEIAAIEBERSA 158
QY 83 G-----HNNPQVMDTE-HSNPDDSGSVPLDPITWGPERRKEESSG 121
DB 159 ARKRSRSPKRGKKKSQENKKNKNIQDIQLTISEPDTSANMRDSEKPEDEKPSA 218
QY 122 HFLVATGTAAAGGGGLIASPKIGATPLPE-STPAPPPPPPPPPPGVSGHINPLI 180
DB 219 SALDEPA-----TLQEVASQEVPELATPPAPAMEQPEPDERLEAAACEVNDL 266
QY 181 LELRLVLOQROIHMOMTEQICROYLLGSLGQTVGAPASPSLPGTASSTKRLPLF 240
```


Db 267 GEE-----EEEEEEEEEE-----DDDDDELEDEGEFEASMP----- 300
QY 241 SPIKPVQTSKTIASSSSSSSSSGAETPKQAFPHLYHPLGSHPSAGVGSRHKPTTAP 300
Db 301 -----NENSVKEPE-----IRDEKPE----- 316
QY 301 SPALPGSTQOLIASPHLAFSTTGLLAQCLGAARGLE--ATASPLGLPKKNGSGELSYG 358
Db 317 -----EDLLEPKTTSSET-----LEDCESEVTPAMQIPIR--TKEEANG 352
QY 359 EVMGPLEKPGRHKCRFCACVFGSDSALOIHLSRH--TGERPYKCVNCGNRRFTTGNILKY 416
Db 353 DVFETTFMFP-----CQCEKERTTKQGLERHMHIIHISTVNHAFKCKYCGKAFGTQIRRR 407
QY 417 HFRHRH--EKYPHVQNNPVRPHLDVYITSSGLPYGMSV-----PP----- 456
Db 408 HERRHAGLKRKPSQTLQIP--SEDL-----ADGKASGENVASKDDSSPPLGPDCLIMN 459
QY 457 -EKAEEEAATPGGVEKRPVLAATTLALATESLTLSTAGTAT--APGLPAFNKFIYM 512
Db 460 SEKASODTI-----NSSVVEENGEVRELHPCKYCKKVGTHTNMRRHORVHERHLIP 512
QY 513 KAVEPKKADENTPPGSEGAISGV----- 537
Db 513 KGVRRKGLLEPPAPQAQNTQNVYVSTEEPEEGEADVDYIMDISNISENLNYIDG 572
QY 538 -----AESSATATLMOLSKLMTSLPSKALLTNHKS----- 567
Db 573 KIOTNNNTSNCVDYIEMESASADLVGINCLLT--PYVEITQNIKTQVPYTEDLPKEPLG 630
QY 568 -----TGSFPL--PICARALGASPSFT-----SK 589
Db 631 STNSEAKKRRRTASPPALPKIAETISDPPVPGCSLSPLSISTTEAVSFHKEKSVYLSK 690
QY 590 LOOLVE-----KIDROGAAVNTSAAGCAPTTA-----PAFSSASSGPN 629
Db 691 LKQLOLQODKLTTPRAGISATEIATKLGVPVCVSAFASMLPVTSRFRKRTSSPPSQHSP- 749
QY 630 QCVYCLAVLSPRALRLHYGQHGGERPFKCKYCGRAFSTRGNL--RAHFYGHASPA-- 684
Db 750 -----ALR--DFGKPSD-----GKAAMDAGLTKSKSKLESHSDSPAMS 786
QY 685 -----ARAONSCPIC-----OKFT-----NAVTLQOHVHMLGQIPINGST----- 721
Db 787 LSGDERETVSPPCDEYKMKSEMTASSAFSSVCNQPLDLSSGVKKAEGTGKTPIVQWE 846
QY 722 -----ALPEGGAGAOENGSEOSTVSGAG-----SF 746
Db 847 SVLDLSVHKHCHSDSEGEKFEKESHSHVQPTCSAVKKRPTTCMLQKVLNEXNGIDLPEVN 906
QY 747 PQOOSQOPSPBEELSEEEBE-----DEEBEDYDEDESLAGSGSESGE--KAIS 795
Db 907 PADGTRSPSPCKSLAQDPDLGPGSGFPAPVVESTPPDVCSSPALOTPLSLSSGOLPIL 966
QY 796 VRGD-----SEASGAEEBVGTVAAATAGKENDSMNK 828
Db 967 IPTPSSPPCPVLTVAITPPPLIPTVPLPAPSSSASPHPCSPPLSNATQOSPILISP 1026
QY 829 TTQOSSLPPPP--PPDSLDQPPOMEGSSCVLGKKEGKPERSS-----PASKLTP-- 879
Db 1027 TVSPSPSPIPRVEPIIMASAPGPPTLSSSSSSSSSSSSSSSSSSSSSPPLSAISSVY 1086
QY 880 -----EBEATSVTLVEELSLQEAM--RKEPGESSSRK-----ACEVCGOAFPS 920
Db 1087 SSGGNLEASLPMISKQELNENGLKPREEPQSAEDVYVQETFNKNFVNCVCSPLFS 1146
QY 921 QAALEEHQKTHPKEGPLFTVCRCRGFLERATLKKHMLAH 961
Db 1147 IKDLTKHLHSIAHEWP--FKCEFCVOLFKDKTDLSEHRLHL 1186
RESULT 21

US-08-399-411-4
; Sequence 4, Application US/08399411
; Patent No. 5831008
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,411
; FILING DATE: 06-MAR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-399-411-4
Query Match 4.2%; Score 222; DB 2; Length 1719;
Best Local Similarity 17.8%; Pred. No. 2,1e-06;
Matches 232; Conservative 142; Mismatches 425; Indels 502; Gaps 53;
QY 9 SRLGVPAEPALG-----GDASEED-----HPQYCAKCAQFTDTE--F 47
Db 40 TRIGVWATKPIILKGGKFGVGDKKRSQYKNNYMEVYYPNL--GWMCIDATDEKGNW 98
QY 48 LAHON-ACS-----TDDPV-----MYIIGOENNNSSASSEPRPE 82
Db 99 LRYVWACSGEENLFPLEINRAIYKTKLPPIAAGELLVWYNGEDNPEIAAIEERAS 158
QY 83 G-----HNNPQVMDTE--HSNPPDSGSSVPTDPTWGPERRGEESG 121
Db 159 ARSKSSSPKSRGKKKSKQENKKNKIQDIQKTSSEDFTSANNMDSAGEPREDEKPSA 218
QY 122 HFLVAATGTAAGGGGLIASPKLGATPLPPE--STPAPPPPPPPPPPVGGSHLNIPLI 180
Db 219 SALEQPA-----TLQEVASQEVPELATPPAPMBPQPPDRLLEAAACEVNDL 266
QY 181 LEELAVLQGRQIHQMOTEDICQVILLGSLGTVGAPASPSSELPGTGTASTKPLPLPPE 240
Db 267 GEE-----EEEEEEDEEBE-----DDDDDELEDEGEFEASMP----- 300
QY 241 SPIKPVQTSKTIASSSSSSSSSGAETPKQAFPHLYHPLGSHPSAGVGSRHKPTTAP 300
Db 301 -----NENSVKEPE-----IRDEKPE----- 316
QY 301 SPALPGSTQOLIASPHLAFSTTGLLAQCLGAARGLE--ATASPLGLPKKNGSGELSYG 358
Db 317 -----EDLLEPKTTSSET-----LEDCESEVTPAMQIPIR--TKEEANG 352
QY 359 EVMGPLEKPGRHKCRFCACVFGSDSALOIHLSRH--TGERPYKCVNCGNRRFTTGNILKY 416

Db 353 DVFETFMPP-----CQHCERKFTTKQGLERHMHIIHISTVNHAFKCKYCGKAFQTQINRR 407
QY 417 HFHRHR---EKYPHVMNHPVPEHLDYVITSSGLPYGMSV-----PP----- 456
Db 408 HERRHAGLKRKPSQTLP-----SEDL-----ADGKASGENVASKDDSSPSLGPDCILMN 459
QY 457 -EKAEEBAATPGGVERKPLVASTALSTATESLITLLSTAGTAT---APGLPAFKFVLM 512
Db 460 SERASODTI-----NSSVVEENGEVKELHPCYCKYKVFQHTNNRRHQRVHERHLIP 512
QY 513 KAVEPKNADEPNPGSEGSALISG----- 537
Db 513 KGVRRKGGLEEOPPAOAOQTQNYVVPSTEEPEEGEADVDYIMDISNISENLNYIDG 572
QY 538 -----AESSTATLMQSLKMTSLPSMALLTNHFKS----- 567
Db 573 KIOTNNNTSNCVDIEMMSASADLGINCLLT--PYVETTONIKTQVPTEDLPKEPLG 630
QY 568 -----TGSFPL-PLCARALGASPSPT-----SK 589
Db 631 STNSEAKKRTASPPALPIKAEFTDSDMPVSCSLPLSISTTEAVSFHKEKSVYLSK 690
QY 590 LQOLVPE-----KIDROGAVAVTSAASGAPTTSA-----FAPSSASGPN 629
Db 691 LKOLLOTQDKLTTPAGISATEIATKAGVCSAPAMLPVTSRFRKRTSSPPSSPOHSP 749
QY 630 QCVICLRVLSCPRALRLHYGOHGEGRPFCKVCGRAFSTRGNL---RAHFVGHKASPA-- 684
Db 750 -----ALR-DFGKPSD-----GKAAMTDALITSKSKSLSHSDSPAWS 786
QY 685 -----ARAONSCPTIC-----QKFT-----NAVTLQOHVRHMLGGOIPNGGT----- 721
Db 787 LSGRDERETVSPCEDEYKMSKEMTASSAFSSVCNOQPLDLSSGVKQAKEGTGKTPVQWE 846
QY 722 -----ALPEGGAOENGSQSTVSGAG-----SF 746
Db 847 SVLDLSVHKHHCSDSEKREKESHVSOPTCSAVYKRRKPTTCMLQVYLLNEYNGIDLPEVN 906
QY 747 POQSOQSPSEELSEEEBE-----DEBEEDVTDEDSLARGSESGE--KAIS 795
Db 907 PADGRSPSPCKSLAEQDPDLPGSGFPAPTVESTPVCSSPALQPTSLSSGQLPPL 966
QY 796 VRGD-----SEBASGAEEVGYVAAAATAGKEMDSNEK 828
Db 967 IPTDSSPPPCPVLTVAAPPPLLPVPLPAPSSSASPBCPSPLSNATAOSPLILSP 1026
QY 829 TTOQSSLPPPP--PPSDLOPOPMEQSSGVLGKKEGKPERSSS-----PASALTP-- 879
Db 1027 TVSPSPSPPIPVPEPLMSASAPGPTLSSSSSSSSSSSSSSSSSSSSSPPLSLATISSVY 1086
QY 880 -----EGEATSVTLVEELSLQELAM--RKEPGESSSRK-----ACEVCGQAFPS 920
Db 1087 SSGDNLLEASLPMSIFKQOELNGLKPREPQSAEQDVVQETFNKNKVFVAVCESPELIS 1146
QY 921 OAALEEHOQTHKREGFLTCVPCROGFLERATLKHMILAH 961
Db 1147 IKDLTKHLISHAEMP-FKCEFCVOLFKDXTDLSEHRLH 1186

RESULT 22
US-08-516-859A-4
Sequence 4, Application US/08516859A
Patent No. 6069231

GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California

/?
/? COUNTRY: USA
/? ZIP: 92122
/?
/? COMPUTER READABLE FORM:
/? MEDIUM TYPE: Floppy disk
/?
/? COMPUTER: IBM PC compatible
/? OPERATING SYSTEM: PC-DOS/MS-DOS
/? SOFTWARE: PatentIn Release #1.0, Version #1.25
/?
/? CURRENT APPLICATION DATA:
/? APPLICATION NUMBER: US/08/516,859A
/? FILING DATE: 18-AUG-1995
/? CLASSIFICATION: 514
/?
/? PRIOR APPLICATION DATA:
/? APPLICATION NUMBER: US 08/399,411
/? FILING DATE: 06-MAR-1995
/?
/? PRIOR APPLICATION DATA:
/? APPLICATION NUMBER: US 08/292,683
/? FILING DATE: 18-AUG-1994
/?
/? ATTORNEY/AGENT INFORMATION:
/? NAME: Campbell, Cathryn A.
/? REGISTRATION NUMBER: 31,815
/?
/? REFERENCE/DOCKET NUMBER: P-LJ 1776
/?
/? TELECOMMUNICATION INFORMATION:
/? TELEPHONE: (619) 535-9001
/?
/? INFORMATION FOR SEQ ID NO: 4:
/? SEQUENCE CHARACTERISTICS:
/? LENGTH: 1719 amino acids
/? TYPE: amino acid
/? TOPOLOGY: linear
/?
/? MOLECULE TYPE: protein
/?
/? US-08-516-859A-4

Query Match 4.2%; Score 222; DB 3; Length 1719;
Best Local Similarity 17.8%; Pred. No. 2,1e-06;

Matches 232; Conservative 142; Mismatches 425; Indels 502; Gaps 53;

QY 9 SRLGVPAEPFAELG-----GDASEED-----HPQYCAKCAQFNDPTE--F 47
Db 40 TRIGVWATKRLKGGKRGPFVGDKKRSQYKNNYIMVEYYIYNL-GMKIDATDPKGNW 98
QY 48 LAHON-ACS-----TDPVY-----MVIIGQENPNNSASSEPRPE 82
Db 99 LRYVMNACSGEONLPLEINRAIYYKTLKPIARCEELLVWNGHDNPRIAAIEBERAS 158
QY 83 G-----HNNPOYMDPE-HSNRPDSGSSVPPDPTMGPERRGESG 121
Db 159 ASKRSKSPKSKKKKSOENKKNKGNKIQIDLKTSEDPFTSANMRDSAGPKDEEEKPSA 218
QY 122 HPLVATGTAAAGGGGULLASPKIGATPLPPE-STPAPPPPPPPPPPGVSGHLNIPLI 180
Db 219 SALEQPA-----TLQEVASQEVPEPLATPAPAMEPQRPDERLEAAACEVNDL 266
QY 181 LEELVLOQROIHOMQTEICROVLLLSIGOTVGAAPASSELPGTGTASTKPLLPLE 240
Db 267 GEE-----EEEEEEDEEEBE-----DDDDDELEDEEEFEASMP----- 300
QY 241 SPIKPVOTSKTLASSSSSSSSSGAETPKQAFPHLYHPHLSQGHSPSAGVGSNHPTRAP 300
Db 301 -----NENSVKEPE-----ICDERK----- 316
QY 301 SPALGSTDOLIASPHLAPPTTGLLAQOCIGARGLE--ATASPLGLPKPKNGSELXSG 358
Db 317 -----EDLLEPRTTSEET-----LEDCEVTPPAMQIPR--TTEEANG 352
QY 359 EVMGLEKPGGHNKCFCKAKVSGDSALQIHLNSH--TGERPYVCNVCNRRPTTRGNLKV 416
Db 353 DVFETFMPP-----CQHCERKFTTKQGLERHMHIIHISTVNHAFKCKYCGKAFQTQINRR 407
QY 417 HFHRHR---EKYPHVMNHPVPEHLDYVITSSGLPYGMSV-----PP----- 456
Db 408 HERRHAGLKRKPSQTLP-----SEDL-----ADGKASGENVASKDDSSPSLGPDCILMN 459
QY 457 -EKAEEBAATPGGVERKPLVASTALSTATESLITLLSTAGTAT---APGLPAFKFVLM 512

```
Db 460 SEKASODTI-----NSSVVEENGVEKELHPCKYCKKVFQHTHTMMRRHQRRVHERHLIP 512
OY 513 KAVEPRKNKADENTPPSEGSATSGV----- 537
Db 513 KQVRRKKGLEEPQPPAEQAQATONVYVSTPEEREGEADDVYIMDISNISENLNTYIDG 572
OY 538 -----AESSTATIMQLSKLMTSLPSSMALLTNHFS----- 567
Db 574 KIQTNNTSNCDVTEHMSASADLYGINCLLT--PVTVEITQNIKTQVPTVEDLPKEPLG 630
OY 568 -----TGSFPL-FLCAALALASPSET-----SK 589
Db 631 STNSEAKKRRRTASPPALPKIKAEFTSDOPMVPSCLSLPLSTTTEAVSPHKEKSVYLISK 690
OY 590 LQOLV-----KIDRGAAVATSAASGAPTTSA-----PAPSSSASSGPN 629
Db 691 LKQLOLTQDKLTTPPAGISAEIATKLGVCVGNAPAMLPVTSRFRKRTSSPPSPQHP- 749
OY 630 QCVICLRVLSCPRALRLHYGQHGGERPEKCKVCGNAFSTRGNL--RAHFVGHKASPA-- 684
Db 750 -----ALR-DFGKRPD-----GKAAMTDAGLTTSKSKLSHSDSPAWS 786
OY 685 -----ARAONSCPIC-----QKFT-----NAVTLQOHVBMHGGQIPNGT----- 721
Db 787 LSGRDERETVSPCFDEYKSKKWTASSAFSSVCNQOPLDLSGVKOKAEGTCKTPVQWE 846
OY 722 -----ALPEGGAQENGSEOSTVSGAG-----SF 746
Db 847 SVYLDLVHKKHCHSDSEKKEFKESHSVOPTCSAVKKRKPPTCMLOKVLYNLNEYNGIDLPVEN 906
OY 747 PQOQSQOPSEBEELSEEEER-----DEEEDVYTDDESLAGRGESGGE-KAIS 795
Db 907 PADGTRSPSPCKSLAOPDPDLPGSGFPAPTVESYPDVCPSSPALQTPSLISSGOLPPL 966
OY 796 VRGD-----SEASGAEEVEGTVAAANAATGKMDSENEK 828
Db 967 IIPDPSSPPCPPLVYATPPRPPLLYPLRPASSASHPICSPLSNATQASPLPLISP 1026
OY 829 TTSQSSLP--PPDSLDOPQPMBOGSSVLGKKEGKPERSSS-----PASALP-- 879
Db 1027 TVSPSPPIPEVPEPLMSAASGPPPLTSSSSSSSSSSSSSSSSSSSSSPSPPLAISSEV 1086
OY 880 -----EGEATSVLVLELSLOEAM--RKEPESSSRK-----ACEVGOQAPPS 920
Db 1087 SSGDNLASLPMISFKQBLENEGKPREEPQSAEODVVQETFNKNFVCNVCESPPLS 1146
OY 921 QAALEHQKTHPKEGPLFTCVCFRGFLERATLKHHMLLAH 961
Db 1147 IKDLTKHLSTAHEMP-FKCEFCVQLFKDKTDLSEHRLFLH 1186

RESULT 23
US-09-586-472-4
: Sequence 4, Application US/09586472
: Patent No. 6323335
GENERAL INFORMATION:
: APPLICANT: Huang, Shi
: TITLE OF INVENTION: Retinoblastoma Protein - Interacting
: Zinc Finger Proteins
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
: ADDRESS: Campbell & Flores LLP
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```

:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/586,472
: FILING DATE: 01-Jun-2000
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 09/528,706
: FILING DATE: 17-MAR-2000
: APPLICATION NUMBER: US 08/516,859
: FILING DATE: 18-AUG-1995
: APPLICATION NUMBER: US 08/399,411
: FILING DATE: 06-MAR-1995
: APPLICATION NUMBER: US 08/292,683
: FILING DATE: 18-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 4130
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1719 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-586-472-4

Query Match 4.2%; Score 222; DB 4; Length 1719;
Best Local Similarity 17.8%; Pred. No. 2,1e-06;
Matches 232; Conservative 142; Mismatches 425; Indels 502; Gaps 53;

OY 9 SRUGVPAGEPRLG-----GDASED-----HPOVCACCAQFTPTPE--F 47
Db 40 TRIGVMAKPKILKGKFGPEVDKRRSQVKNVYMEVYVPL-GWACIDATDPEKGM 98
OY 48 LAHON-ACS-----TDPV-----MYLIGQENPNNSASSEPPRE 82
Db 99 LRYVNMACSGEONLFPLEINRAIYKTKLPARGELLWYNGEDNPEIAAIEEERAS 158
OY 83 G-----HNPQVMDTE-HSNPPDSSSVFTDPTWGPERKGEESG 121
Db 159 ARSKRSPSKRKKKKQCNKNKGNKIDQILKTSEDDFTSANNRDSAEGRKEDEKPSA 218
OY 122 HFLVAATGAAGGGGLIASPKLGATPLPE-STAPRPPPPPPPPPGVSGHLNPLI 180
Db 219 SALRQPA-----TLOEVASQEVPELATPAPAMEPOPEPDERLEAAACEVNDL 266
OY 181 LEELRVLOQRTHQMTQEOICRQVLLGSLQTVGAPASPSELPGTASTKPLPLF 240
Db 267 GEE-----EEEEEDEDEEE-----DDDDDELEDEGEESASMP----- 300
OY 241 SPIKPVQTSKTIASSSSSSSSSGAETPKQAFHLHYPLSGHPFSAGVGRSHKTPAP 300
Db 301 -----NENSVKEP-----IRCDERP----- 316
OY 301 SPALPGSDQLASPLAPLPSTGTGLAAGCLGARGLE--ATASGLLKPKNGSELSTYG 358
Db 317 -----EDLLEPFTTSEET-----LEDCEVTPAMQILP--TKEEANG 352
OY 359 EVMGPLEKPGGRHKCFKCAKVFGSDSALOILNRSH--TGERPYKCNVCGNRRFTTGNLKV 416
Db 353 DVEETFMFP-----COHCERKFTTKQGLERHMHISTVNHAFKCKYCGKAFGTQINRRR 407
OY 417 HFHRR-----EKYPHVQMNPRVPEHLDYVITSSGLPYMSV-----PR----- 456
Db 408 HERRHEAGLKRKRSQTLQ--SEDL-----ADKASGEVNAASKDSSPSPGLPDLINN 459
OY 457 -EKAEEEAATPGGGERKPLVASTTALSAATESLTLSTASGAT--APGLPAFNKFLVM 512
Db 460 SEKASODTI-----NSSVVEENGVEKELHPCKYCKKVFQHTHTMMRRHQRRVHERHLIP 512
```

513 KAVEPKKADENTPPGSESAISGV----- 537
513 KAVRRKGLLEPPQPAEQATQNYVVPSTEPEEGEADVYIMDISNSENLYIDG 572
538 -----AESSTATMOLSKIMTSLPMSWALLTNHKS----- 567
573 KIOTNNNTSNCVDIEMESASADLYGINCLLT--PYVEITQNIKTQVPYTEDLPKEPIG 630
568 -----TGSFPL-PLCARALGASPSET-----SK 589
631 STNSEAKKRRRTASPPALPKIAEDTSDPMVPSCLSLPSTISTTEAVSFHKEKSVYLSK 690
590 LQOLV-----KIDRGAAVAVTASAGAPTTA-----PAPSSASSGPN 629
691 LQOLQOTODKLTTPAGISATRIAKIGPYCVSAPASMLPVTSSRFKRRRTSSPPSPQHP 749
630 OCVCILRVLSCPRALRLHYGGHGERPPKCKVCGRAFSTRGNL--RAHVGHKASPA-- 684
750 -----ALR-DFGKRPD-----GRAAWTDAGLTSSKSKLSHSDSPAWS 786
685 -----ARAONSCPTC-----QKFT-----NAVTLQOHVRMHLGGQIPNGGT----- 721
787 LSGRDERETVSPCFDEYKMSKEWTFASSAFSSVCNQOPLDLSGVAKAGTGTPTVQME 846
722 -----ALPBGGAOENGSEOSTVSGAG-----SF 746
847 SVLDLSVHKHCHSDSEKEREKESHVQPTCSAVKKRPTTCLQKVLNENYNGIDLPEVN 906
747 PQOQSOQPSPEEELSEEEEEE-----DEEEEDVTDDSLAGKRESGGE-KAIS 795
907 PADGTRSPSPCKLEAOPDPLGSGFPAPTVESTPDVCPSSPALQTPSLSSGOLPPL 966
796 VRGD-----SEASGAEEEVGYAAATAGKEMSNK 828
967 IPTDSSPPCPVLTVAATPPPLLPVPLPAPSSASAPHPCPSPLSNMTAQSPILSP 1026
829 TTQOSSLPPPP--PPSLODOPQMEQSSGVLGKKEGKPRRSS-----PASALTP-- 879
1027 TVSPSPSLPPEPLMSAASPPPTLSSSSSSSSSSSSSSSSSSSSSSSPPLSAISSV 1086
880 -----EGEATSVTLVETLSQEM--RKEGESSSRK-----ACEVCQGAFPS 920
1087 SSGDMLASLPMISFQOELENGIKLREPOSAAEQDVVQETPNKNVCVNCESPFLS 1146
921 QALDEHQTHKEGFLFTVFCRCQGLERATLKKHMLAH 961
1147 IKDLTKHLSTHAEMP-FKCEFCVOLFKDKTDLSEHRELLH 1186

RESULT 24

US-09-528-706-4

Sequence 4, Application US/09528706

Patent No. 6468985

GENERAL INFORMATION:

APPLICANT: Huang, Shi

TITLE OF INVENTION: Retinoblastoma Protein - Interacting

TITLE OF INVENTION: Zinc Finger Proteins

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell & Flores LLP

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/528,706

FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/516,859
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1776
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-528-706-4

Query Match 4.2%; Score 222; DB 4; Length 1719;

Best Local Similarity 17.8%; Pred. No. 2.1e-06;
Matches 232; Conservativity 142; Mismatches 425; Indels 502; Gaps 53;

9 SRLGVAPGEPAELG-----GDASED-----HFQYCAKCAQFTDPTPE--F 47
40 TRIGVWATKPLILGKKRPPEYGDKKRSQVKNVYMEVYVNTL-GWMCIDATDPEKGNV 98
48 LAHQN-ACS-----TDPPV-----WVITIGGOENPNNSASSEPRE 82
99 LRYVNMACSGEEDNLPLEINRAITYKTLKPLAPGBELLWYNGENPELAAIIEERAS 158
83 G-----HNNQVMDTE-HSNPPDGSVVPTDPTWGERGESSG 121
159 ARSKRSPPSKRKKKKSOENKKNKGIODIOLKTSEPDFTSANMRDSAGPKDEEKPSA 218
122 HFLVATGTVAGGGGGLILASPKIGATPLPPE-SPPAPPPPPPPPPPPVGSGLNIPLI 180
219 SALEQQA-----TLQEVASQEVPELATPAPAMEPQPEDEPLEAACEVNDL 266
181 DELRLVLOQROIHQOMQTEQICRQVLLGSLQVYGAPAPSELPQTGTASTTKPPLPLP 240
267 GEE-----EEEEEEDEEE-----DDDDDELDEGEESAMP----- 300
241 SPIKPYQTKTILASSSSSSSSSGAETPKQAFPHLYHFLGSOHPFSAGVGSNKPFPAP 300
301 -----NENSVKKEPE-----IRCKDEK----- 316
301 SPALPGSTDLIASPHLAFPTTGLILAAQCIGAARGLE--ATASPLGLPKNGSGELSYG 358
317 -----EDLLEEPRTTSET-----LEDCEVPTAMQIPLR--TKREANG 352
359 EVMGPLEKGGGRHKCFCAKAVFGSDALQIHLRSH--TGERPYKCNVCGNRTTTRGNLY 416
353 DVFETTFMP-----CONCEKFTTKQGLERHMHISTVNAFAFKCKYCKAGKAGTQINRRR 407
417 HFHRRH--EKYPHYOMNPVPVEHLDYVITSSGLPYGMSV-----PP----- 456
408 HERRHAGIKRRPSQLP-----SEDL-----ADGKASGENAVASKDDSSPSPISGPDCLIN 459
457 -EKAEEAATPGGVERKPLVASTALSTATSLTSLTSTAGTAT--APGLPAFKFVLM 512
460 SEKASQDTI-----NSSVVEENGVEYKELHPCKYCKKVFSTHNMRRHQRVHRHRLIP 512
513 KAVEPKKADENTPPGSESAISGV----- 537
513 KGVRRKGLLEPPQPAEQATQNYVVPSTEPEEGEADVYIMDISNSENLYIDG 572
538 -----AESSTATMOLSKIMTSLPMSWALLTNHKS----- 567
573 KIOTNNNTSNCVDIEMESASADLYGINCLLT--PYVEITQNIKTQVPYTEDLPKEPIG 630

QY 568 -----TGSEPL-PLCARALGASPEET-----SK 589
Db 631 STNSEAKKRTTASPPALPKIAETDSDPMVPSCSLSLPLSISTTEAVSFHKEKSVLYLSK 690
QY 590 LQOLIVE-----KIDROGAVAVDSAAAGCAPTTS- ----PAPSSASSGPN 629
Db 691 LKQLOLOQDKLTPPAGISATEIAKLGVCVSAFASMLPYTSSFFKRTSSPPSPQHS- 749
QY 630 QCVICLRYLSCPRALRLHYGQHGGERPEKCKVCYGRAFSYRGNL- --RAHFVGHKASPA- 684
Db 750 -----ALR-DFGKPSD-----GKAAMTDAGLTKSKSKLESKSDSPAWS 786
QY 685 ----ARAQNSCPIC-----QKFT-----NAVTLQOHVRLHGGQIPNCGT----- 721
Db 787 LSGRDERETVSPCFDEYKMSKEWTASSAFSSVCNQOPLDSSGVKOKABGTGKTPVQWE 846
QY 722 -----ALPEGGAOENGSEOSTVSGAG-----SF 746
Db 847 SYVLDLVYHKHCKSDEGKEFESHSHVQPTCSAVKKRPTTCMLQKVLNEXNGIDLPEVN 906
QY 747 PQOQSOQPSPEELSEEEEEE-----DEEBEDVTDEDSLGRGSESGE-KAIS 795
Db 907 PADGTRSPSPCKSLAEOADPDGLGPGSGFPAPYVESTPDVCPSSPALQTPSLSGQLPPL 966
QY 796 VRGD-----SEASGAEEVGTVAATAAGKEMDSNEK 828
Db 967 IPIPDSPSPCPVLYVATPPPLPLPYVLPAPSSASPHPCSPUSNAVAQSPPLPILSP 1026
QY 829 TTOQSSLP-PPDSLDOPQPMEGSSGVLGKKEGKPERSSS-----PASALTP- 879
Db 1027 TVSPSPPIPTVEVPLMAKSAQCPRLSSSSSSSSSSSSSSSSSSSSSPSPPLAISIV 1086
QY 880 -----EGEATSVTLVEELSLOEAM--RKEPGESSSRK-----ACEVCGQAFPS 920
Db 1087 SSGNDLNASLPMISFKQBELENEGLKPREEPOSAAEQDVVVOETFNNKFNVCNVESPPLS 1146
QY 921 QALLEEHOQKTHPKEGPLFTVCYCRQGFLEKRLATLKHMHLAH 961
Db 1147 IKDLTKHLSTHAEMP-FKCEPCVOLFKDKTDLSEHFLHL 1186

RESULT 25
US-09-262-773-2
; Sequence 2, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: MY143
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 648
; TYPE: PRT
; ORGANISM: human
US-09-262-773-2

Query Match 4.2%; Score 219.5; DB 4; Length 648;
Best Local Similarity 19.5%; Pred. No. 8.8e-07;
Matches 136; Conservative 65; Mismatches 263; Indels 233; Gaps 20;
QY 60 VMVITGGENPNSSASBRPRGCHNNPQVMDTHSNPRDSSGVPTDPTWGERGEES 119
Db 135 VIVHVGQGVLSSEETVNLGAEPSPN-----ELQDVQVST-----PQSPPEET 178
QY 120 SGHFLVAATGTGAAGGGGLASPKLGATPLPRESTP-----APPPPPPPPP 167

Db 179 T-----QSPDLGA---PAEQRPQHEELQTLQSESEVPVEDPDLF 215
QY 168 PGVSGHLNIPLLEELRY-----LQORQHOMQTEQICQVL 206
Db 216 AERSSGDSSEWALLTALSOGLVTFKDVAVCFSDQSDLDLPTQKEFYGEVLEDCGIVV 275
QY 207 LLSLQGTVCAPASPSELPGTASTKPLPLFSPRIKPVQTSKTLAASSSSSSSSSSGAE 266
Db 276 ---SLSPPIPPDELISOY-----REEEPWVDIQEPQTEPEILSFYTGDKSKDEE 326
QY 267 TPQAFPHLYPHLSQHPFSAGVGRSHKPTAPASPALPGSTQDLIASPHLAFSTTGL 326
Db 327 CLEQEDLSLEB-----IHRPV-LGEDEIHQTPDM-----EIVFEDNPGRL 365
QY 327 AAOCLGAARGLLEATASPEGLKPKNGSGELSYGEVMPLEKPGGRHRCFCAYKVGSDAL 386
Db 366 NERRFQTN-----ISOVNSFVNLRETTPVHPL--LGRHHDCSVGCKSFTCNHSL 412
QY 387 QIHRSHTGERPYKCNVCGNRRFTTGNLKVHFHREKYPHVOYMNPHVPEHLDVITSS 446
Db 413 VRHLRTHGKPYKCMCEGKSTYTRSHLARHOKVHKMNPY----- 453
QY 447 GLPYGMSVPRPKAEDEAATPGGVERKPLVASTTALSAATESLTLLSTAGTATAPGLPAF 506
Db 454 -----KYLNRKNLEETSPVTOAERTP----- 475
QY 507 NKEYLMKAVEERKNKADENTPPSESGAISGVASSTRATLMQSKLMT-----S 554
Db 476 -----SVERPYRCD-----CGKHFRMTSDVLRHOKTHHGEKPFCTIGKS 517
QY 555 LPSWALLNHFKSTGSFPLPLCARALGASPEFSKIQLOLVEKIDROGAVAVTSAAGAPT 614
Db 518 FSQKSVLTHQRI-----HIGKPYLQGECEDESEHRYLAHRTTHAAEEL- 564
QY 615 TSADAPSSASSGPNOCVICTLRYLSCPRALRLHYGQHGGERPEKCKVCYGRAFSYRGNLRA 674
Db 565 -----YLSECGRCFTHSAAFAKHLRGHASVPRCNCNEGKFSFRD- 606
QY 675 HFVGHKASPARAQAONSCPIQCKKFTNAVTLQOHVRMH 711
Db 607 HLVRHQRTHGKRPCTCGKSFSGYHLIRHQRTH 643

RESULT 26
US-09-262-773-4
; Sequence 4, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: MY143
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 640
; TYPE: PRT
; ORGANISM: human
US-09-262-773-4

Query Match 4.1%; Score 218.5; DB 4; Length 640;
Best Local Similarity 20.7%; Pred. No. 1e-06;
Matches 129; Conservative 61; Mismatches 246; Indels 187; Gaps 19;
QY 145 LGATPLPPE--STPAPPPPPPPPPPGVSGHLNIPLLEELRYLQORQHOMQTEQIC 202
Db 144 LGAEPSPNELQDPVOSTPQSPPEETQSPDLGAP---AEQRPQHEELQTLQESE--- 197

```
OY 203 ROVLLLSIGQVGPASPSELPGTGTASTKPLPLFSPDKPYQTSKTLASSSSSSS 262
Db 198 -----VPPEDP-DLPARESSGDSEWALLTALSQGLVTRDAVVCPSQDQMS 244
OY 263 SGAEPTKPAEFHLVHPLGSOHPFSAGVGRSHKPTAP-----SPALPGSTD-OLI 312
Db 245 DLDPTQKE--FYGEVLEBE---DCGIYVLSLFPPIPRPDEISQVREDEPWPDQEPET 298
OY 313 ASPHLARPTSTGLA---AQCGIAA-RGLEATASPGLKP---KNSGELSYGEVMPLE 365
Db 299 QEPFLISFTYIGDRSKDEECLQEDLSLEDIHRYPLGEPEIHQTPMWEIVFEENPGLN 358
OY 366 KP-----GGRHKCRFCAKVFPSPDALQIHRSHNGERYK 400
Db 359 ERREPTNISQVNSFVNLRETTTPVHPLLGRHHCQVCGKSTCSHLYRLKHTGEEKY 418
OY 401 CNVGNRPETTRGNLKVHFRHREKYPHOMNPHVPEHLDYVITSSGLPYGMSVPERKAE 460
Db 419 CMBCGKXYTRSSHLARHQVKHKNAPY-----KYPILRNKN 453
OY 461 EEAATGCGGERKPLVASTTALSATESLTLSTASGTATAPGLPAPNKVFLMKVPERKN 520
Db 454 LEETSTPTQAEKRP-----SVEKPYR 474
OY 521 ADENTEPSSGSAISGVAESSTATLMOLSKLMT-----SLPSWALLTNHFKST 568
Db 475 CDD-----CGKHFRTWSDLVHRQRTHTGKPFCTICKGSPSQSVLTHORI- 522
OY 569 GSPPLPLCARALASPETSRLQOLVEKIDRGAVAVTSAASAPPTTAPAPSSASSGP 628
Db 523 -----HLGKPYLGCGEDEFSHRRLAHRKTHAAEEL----- 556
OY 629 NOCVIDLRYLSCPRLRLHYOHGGERPFCKVCGRAFTSGRLRAHFVGHKASPARAQ 688
Db 557 YLCSGRCRFTTHSAFAKHLGHASVRCRCNECKSPFSRRD---HLVHQHTHGEKP 612
OY 689 NSCPTQCKRTNAVTLQOHVYRM 711
Db 613 FTCPCTGKSPFSRGVHLTRHQRTH 635

RESULT 27
US-08-102-942A-4
; Sequence 4, Application US/08102942A
; Patent No. 5726288
; GENERAL INFORMATION:
; APPLICANT: Call, Katherine M.
; APPLICANT: Glaser, Thomas M.
; APPLICANT: Ito, Caryn Y.
; APPLICANT: Buckler, Alan J.
; APPLICANT: Pelletier, Jerry
; APPLICANT: Haber, Daniel A.
; APPLICANT: Rose, Elise A.
; APPLICANT: Housman, David E.
; APPLICANT: Bruening, Wendy
; APPLICANT: Darveau, Andre
; TITLE OF INVENTION: Localization and Characterization of the
; TITLE OF INVENTION: Wilms' Tumor Gene
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/102,942A
; FILING DATE: 02-AUG-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-5194A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-102-942A-4

Query Match 4.1%; Score 215.5; DB 1; Length 449;
Best Local Similarity 25.8%; Pred. No. 1e-06;
Matches 101; Conservative 43; Mismatches 130; Indels 117; Gaps 23;

OY 124 LVATGTAGGGGGLILASPKLGA---TP-----PEST-----PAPPPPPPPPG 169
Db 11 LPLAVSSLGGGGGGGL--PVSGARQWAPVLDFAPPGASAYSLGFPAPPPPPPPPP 67
OY 170 VGSGLNIPLELRYLQORQIHQMOMTEQICROYLLLSLQ---TVGA-----PA 219
Db 68 -----PSHFIKQEPSSWGAEPHEQCLSAFTLHFSQFTGTACACRXPGRPP 115
OY 220 SPSEL-PGTGTAASKTPLLPFSPKPY---OTSKL-----ASSSSSSSSSGAETPKQA 271
Db 116 PPGQASSGQARMFNAPNAPYLPSCLESQPTIRNOGSTVITDGAISYGHPSHHAQFPNHS 175
OY 272 FPHLYHPLSGOHPFSAGVGRSHKPTP-----APSPALPGSTDOLIASHPLA---FPS 321
Db 176 FKH-EDPMGOQ---GSLGQOYSVPPVYGCNHPPTDCTGQALLLRPPSSDNLQYM 229
OY 322 TTGLLAQOC-----IGAA-RGLEATASPGLL---KKNSSGELS----- 356
Db 230 TSQI---ECMTWNOMNLGATILKGMAAGSSSVKWTGEGSNHGTYSENHTAPILCAQY 286
OY 357 ---YGEVWG--PLEKPG-----RHKCRF--CAKVFSGDSALQIHLRS 392
Db 287 RIHTGVRGRIQDVRRVSGVAPFLVRSASFSSEKRPFKATYGCCKNRYKLSHLQHSRK 346
OY 393 HGERPYKAV---CGNFTTRGNLKVHPRH 421
Db 347 HTEKPYQCDPKDCERRFRSRDQLKRHQRH 377

RESULT 28
US-09-037-179B-4
; Sequence 4, Application US/09037179B
; Patent No. 6316599
; GENERAL INFORMATION:
; APPLICANT: Call, Katherine M.
; APPLICANT: Glaser, Thomas M.
; APPLICANT: Ito, Caryn Y.
; APPLICANT: Buckler, Alan J.
; APPLICANT: Pelletier, Jerry
; APPLICANT: Haber, Daniel A.
; APPLICANT: Rose, Elise A.
; APPLICANT: Housman, David E.
; APPLICANT: Bruening, Wendy
; APPLICANT: Darveau, Andre
; TITLE OF INVENTION: Localization and Characterization of the
; TITLE OF INVENTION: Wilms' Tumor Gene
; FILE REFERENCE: 0050.1312-011
; CURRENT APPLICATION NUMBER: US/09/037,179B
; CURRENT FILING DATE: 1998-03-09
; PRIOR APPLICATION NUMBER: US 08/102,942
; PRIOR FILING DATE: 1993-08-02
```

```

: PRIOR APPLICATION NUMBER: US 07/614,161
: PRIOR FILING DATE: 1990-11-13
: PRIOR APPLICATION NUMBER: US 07/435,780
: PRIOR FILING DATE: 1989-11-13
: PRIOR APPLICATION NUMBER: US 07/795,323
: PRIOR FILING DATE: 1994-09-27
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 449
: TYPE: PRT
: ORGANISM: Unknown
: FEATURE:
: OTHER INFORMATION: Mutline
US-09-037-179B-4

Query Match      4.1%  Score 215.5; DB 4; Length 449;
Best Local Similarity 25.8%  Pred. No. 1e-06;
Matches 101; Conservative 43; Mismatches 130; Indels 117; Gaps 23;

QY 124 LVANATGACGGGGLLIASPKLGA---TPL-----PEST-----PAPPPPPPPPPG 169
    | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11 LPLAVSSLGSGGGGCGCL--PVSGARQMAPVLDPAPGASVYGLSGPAPPPPPPPPP 67
    | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 170 VSGSHNLPILLELRYLQQRQIHQMOMTEQICRQVLLGSLGQ---TVGA-----PA 219
    | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 68 -----PHSFITQEPDSWGGAEHPHEEQCLAFTHLFGSQFGTAGACRYGFGPGR 115
    | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 220 SPEEL-PTGTASTTKPLPLPSPKIPV---QTSKTL-----ASSSSSSSSSGAETPKQA 271
    | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 116 PPSQASSGQARMFENAPRLPSCLESQPTIRNOGYSTVTEGASVYGHTPSHNAAQFPNHS 175
    | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 272 FPHLYHPLGSOHPFSAGVGSRSHKPTP-----APSPALPGSDQLIASPHLA---FPS 321
    | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 176 FKH-EDMGGQ-----GSLGEQYVSPRPVYGCHTPTDCTSGALLLRTPYSSDNLXQM 229
    | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 322 TTGLLAAQC-----LGA-RGLENTASPGLL-----KPNNGSGLS----- 356
    | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 230 TSQL---ECMTWNOMNLCATLTLGMAAGSSSSSVKWTGQSNHGTEGSEENHTAPILGAAQY 286
    | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 357 -----YGEVMG--PLEKPG-----RHKCRF--CAKYFGSDSALQHLRS 392
    | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 287 RIHTHGVPFRGLQDVARVGVAPTLVRSASSETSEKRRPMCAIPGCNKKRYFKLSHLMHSRK 346
    | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 393 HTGERPYKCNV--CGNRETTTGNLKVHFRH 421
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 347 HTGEKPYQCDPKDCERFRFSRSDQLKHOGRH 377
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 29
US-08-317-310A-15
: Sequence 15, Application US/0317310A
: Patent No. 5858701
: GENERAL INFORMATION:
: APPLICANT: WHITE, MORRIS R.
: APPLICANT: SUN, XIAO JIAN
: APPLICANT: PIERCE, JACALYN H.
: TITLE OF INVENTION: THE IRS FAMILY OF GENES
: NUMBER OF SEQUENCES: 64
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 28 State Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/317,310A

```

```

FLINGDATE: 03-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: JDP-022
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1234 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-317-310A-15

Query Match 4.1%; Score 215; DB 2; Length 1234;
Best Local Similarity 20.4%; Pred. No. 4e-06;
Matches 216; Conservative 111; Mismatches 358; Indels 372; Gaps 55;

QY 2 AHSERSRLVPGERPELGDASEEDHPQVCAKCAQFTDTEFLAHQNA--CSNDPP 59
Db 351 AHHNRGSSRLHPPLNHSRI-----PMDSSRCSPATSVLSSSSTSGCGSTSDC 401
QY 60 VMTYIGQENPNNGSSASSEPRPEGHNNPOVMDTEH--SNPPDSSGSS----VPTDPTWGER 114
Db 402 LF-----PRRSASAVSGSPS--DGGFTSSDEYSSPCDFRFSFVSVPDLSGHPPPA 451
QY 115 RGEES-----SGHFLVAATGTAAGGGGGLLASKLQATPLRP----- 152
Db 452 RGEELSLNYICMGKGASTLAAPNGHYII-----SRGNGHRYIIGCANLITSPALPGDEA 506
QY 153 -----ESTRAPPPP-----PRPPPGVSGS----- 173
Db 507 AGAADLNRFRKRTHSAGTSPTISQKTRPSOSSVASIEETEMRAIYPPROGSGGRLPG 566
QY 174 -----HLNPLILELRVLOQRIHOMQTEQICROVLLGSLGQTVGADAPSEL 224
Db 567 YRHSFAFPTH-SYPEEGLEMHNLERRGCHNRPTSLNLTDD--GYPMMSPGVAPVPSNR 622
QY 225 PGRTASTSKLLP-----LFSPIK--PYQT-----SKTLASSSSSS 260
Db 623 KGNQ--DYPMSPKSVSAPROQIINFRHPQVDPNGYMMSPSGSCSPDIGGSSSS 679
QY 261 SSSGAEPRKQAFHLYNPLGSOH--PFSAGVGRSHKPTRPAPSPALPGSTDLIASPHLA 318
Db 680 SISNA-----PSGSSYGRKWT--NGVGHH-----THALPHAKRPVESGGKL 720
QY 319 PPSTTGLIAOCLGARGLEATASRGLKPKNGSGELSYEVMGRPLEKP----- 367
Db 721 LPTGQDYMNMSPVGDSN--TSSP-----SECYUPR-EDPQHKPVLSTYSYLRPSF 766
QY 368 -----GGRKRCRCAKVFGSDSALOILHRSHTGERPKCNV----- 403
Db 767 KHTQRPGEPCARQ-----HLNLSSSGSLRYATATAEDSSSTSSDLSIG 812
QY 404 --CGNRFTRGNLKVHFRHREKYPHVQMNPRPVEHLDVI-----TSSGLP 449
Db 813 GGYCGAR--PESSL-THPHNN-----VLDPH-LPKKVUTAOTNSRLARTPLRLSLDPP 861
QY 450 YGMSVPRKAE-----EAATPRG-----GGERKPLVASTALSTESTLILS 492
Db 862 KASTLPVREQOQOOSSTLHPRPKSPGEVYNIIEFGSGOGCYLAGPATSRSSPS----VR 917
QY 493 TSAGTATAPGLPAFNKFKVLMKAVERPKNKADENTPRGSGSAISGVAESSTAILMQLSKM 552
Db 918 CRPQHLPARETSGSEYVNMIDGPRBRATWQESGVLEGRl-GRAPGSAIVCPRT-- 974
QY 553 TSLRSMALLTNFKSTGSPRLP--CARALGASSTENSKLOQYLEKIDROGVAVYATSSA 610
Db 975 -SVNP-----SRGDVWTMOIGCPR--QSTVDTSPVAP--VSAYAMRGIAAEKASL 1020

```

QY	611	GAFTTSAPAPSSSSAGCPMOCVTCILRVLSCPRLRLHYGH----	GGERPCKKVCGR	665
Db	1021	PRPTGAAPPSSSSVASSAS-----VPCQATACQATHSSLLGGPOG----	GGMS	10666
QY	666	FSTGRLRLAHFVGHKRPASPARAONSPCICKKFTFNATVTLQOHVHMLGGQIPNGTALPE	725	
Db	1067	APTYRLSPNH--NQSAKYIRLATQG--CRRHRSSE--TFSAPIR--AGNYVPFAGAAVG	1119	
QY	726	GGGAADENSSEOSTVSGAGSFPQOOSOPSPBEELSEEEEEDEEEDVDYDDESLAGRG	785	
Db	1120	GGGGGGGGGGSEVVKRRHSSAF-----ENVMLPR	1147	
QY	786	SESGG--EKAISVRCGSEEAASGAEEFVTAALATAGKEND-----SNEKTTQOOSLPP	838	
Db	1148	GLDGGVSKRSAPVCG--AAGGLEKRLNTI-----DDLAKHEQSDCPQQQSLPP	1196	
QY	839	PPPDSLDOPQPMEGSSGVLGKKECGKPERSSSPAS	875	
Db	1197	PP-----HQPL-----GSNEGNSPKRSSSEDL	1218	

RESULT 30
PCT-US95-13041-15

QY	60	YMYIIGQENPNNSASSERPERBEHNPNQVMDTEH-SNPPSGSS-----VPDPDTWGP	114
Db	402	LF-----PRSSASVSGSPS--DGGFISDSEGGSSPCFRSSFRSVTPPSLGHTPPA	451
QY	115	RGEES-----SGHFVLVATGTAAAGCGGGLIILASPKIGATPLP----	152
Db	452	RGEELSNYICMGKGASTLAAPNGHYIL-----SRGNGNHRIYIPGANIGTSPALPGDEA	506
QY	153	-----ESTPAPPPP-----PPPPPGVSGS-----	173
Db	507	ACAADLDNFRKKTTHSAGTSPPTISHQTPSOSVASIIEEYTEMMAAPVPGGSGCGRLPG	566
QY	174	-----HNTPLLEERVLQOQRIMQOMTEQICRQVLLGLSLQGYAPASPSL	224
Db	567	YRHSATVPPT-SYPEEGLEMHILERRCGHHRPDLSNLHTDD-----GYMMSPEVAVPNSR	622
QY	225	PGGTASSTRKPLP-----LFSPK--PVQ-----SKLASSSSSS	260
Db	623	KGGG-----DYMPMSPKVSADPOQITNPRHRORDPNMGMMSPSGSCSPDIGAGSSSS	679
QY	261	SSSGAETPKQAFELHLYPLGSOH--PFSAGVGSRKHPPTAPSPALPCTSDLIASPHLA	318
Db	680	SLSAA-----PSSGSYGKPMPT-NGVGHH-----THALPAKPPVEGGGGL	720
QY	319	PFTSTLLAQCIGAARGLELATSPGLIKRNSGELSYGVMGPLEP-----	367
Db	721	LPCTGDYMNKSPVGDGN-----TSSP-----SECYYG-EDPQHKPVLSYSLPSRF	766
QY	368	-----GRRHRCRFCAKVFSGDSALQIHLSRHTGERPYKNV-----	403
Db	767	KHTQRPGEDEGARHQ-----HLRLSSSSGRRLRYATATADSSSTSSDSL	812
QY	404	---CGRRFTTRGKLKXHHHREKYPVQNMNPVPEHLDYV-----TSSGLP	449
Db	813	GGYCGAR--PESSL-THPHH-----VLQPH-LPRKPYDAQNTSLAPTRLSTLDP	861
QY	450	YMSVPERKAE-----EAATPG-----GVERKPLVASTALATESITLIS	492
Db	862	KASTLEPRVEEQOQQOOSLHPRPKSPGEVNIIEFGSGQPYLAGPATRSRSPS---VR	917
QY	493	TSAGTATACGLPAFNKFLMKAVEPKKADENTPPGSEGSASISVAESSTATLMQSLKM	552
Db	918	CPPOLHPARETSGSEYNNMMDGPRKRAVMOESGVELGRI-GRAPGSATVCPRTR--	974
QY	553	TSLPSEWALLTNHFKSTGSEPLP--CARALGASPESTKLOOLEKIDROGAVANTSAAS	610
Db	975	-SVPN-----SHGDYMTMGIGCPR--QSVYDISVAPR-VSYADMKGIGIAEKSL	1022
QY	611	CAPTTSAPAPSSASSAGPNOCVICLRYLSCPRALRLHYQH-----GGERPFCKVCGRA	665
Db	1021	PRPTAAPPSTSTASSAS-----VPOCATABOATHSSLLGPOGP-----GGM	1066
QY	666	FSTRGNLRHFHFGHKSPAPARAONSPICOKKFTTNAVTLQOHVMMHGGQIPNGGTALPE	725
Db	1067	AFTRYNLSPNH--NQSAKYIRADTOG--CRRRISSE-TFSAPTR--AGNYTPFGAGAIVG	1119
QY	726	GGGAQENSGEOSTYSGAGSPQOQSOQPSPEBELSHEEBEDEEBEDVDTEDSLARG	785
Db	1120	GGGGGGGGSEDEVKRRSSASF-----ENWMLPR	1147
QY	786	SESGG--EKAIIVRGSSEBASGEEEVGYAAATAAGKEMD-----SNEKTTQOOSLPP	838
Db	1148	GDLGVSKESAVYCG--AAGGLEKSLNTI-----DIDLAKHNSQDCPSQOOSLPP	1196
QY	839	PPPSLDLPQPMEGSSGVLGKKEGGRKERRSSPAS	875
Db	1197	PP-----HQPL-----GSNEENSRRSRSEDSL	1218

RESULT 31
US-08-234-783-4
; Sequence 4, Application US/08234783
; Patent No. 5622835

GENERAL INFORMATION:
APPLICANT: Herlyn, Meenhard
APPLICANT: Morris, Jennifer
APPLICANT: Rauscher III, Frank J.
APPLICANT: Rodeck, Ulrich
TITLE OF INVENTION: WTI Monoclonal Antibodies and Methods of
TITLE OF INVENTION: Use Therefor
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/234,783
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST480USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-234-783-4

Query Match 4.1%; Score 214.5; DB 1; Length 429;
Best Local Similarity 24.3%; Pred. No. 1.1e-06;
Matches 91; Conservative 43; Mismatches 122; Indels 119; Gaps 18;

133 GGGGGLIASPKLGATP-----LPPEST-----PAPPPPPPPPPGVSGHLNIP 178
19 GGGGGCAL--PVSGAQMVAVLDFAPRGASAYGSLGAPRPPRPPRPHS----- 70
179 LLEELRYLQROQHOMOMTEOICROVLLGSLGQ--TVGA-----PASPSEL-PGT 227
71 -----FIKQPSWGAERHEQCLSAFTVHFGQFTGTAGACRYGPRGPPRPSQASSGQ 124
228 GTASTKPLPLFSPRIKRYQVTSKTLASSSSSSSSSSG-----AETPKQAFHLYHPLG 280
125 ARMFPMAPYLPSCLSEOPAIRNOGYSTVTFDGTSPSYGTHPSHNAQFPHNSFKH-EDPMG 183
281 SQHPFSAGVGSRHKPTP-----APSPALPGSTDLIASPHLA---FPSTTGL----- 325
184 QQ-----GSLGEQYQSVPRPVYGCNTPDSCSTGSQALLKRTPRYSSDNLXQMTSOLECMTW 238
326 ----LAAOCLGAARGLEAT-----ASPGILKPKN 350
239 NQMNLGATLKGHSTGVESDNHTTPIILCGAQRHHTGVRGIDQVRRVGVAPTLVRSAS 298
351 GSGELSYGEVMGLKPRGGRHKCRF--CAKYFGSDSALOIHLSHTGERPYKCNV--CGN 406
299 ETSE-----KRP---FMCAYPGCKNRKYFKLSHLQMSRKHTGKPYQCDKDCER 345
407 RFTTGNLKAHFHRH 421
346 RFSRDQLKRHRH 360

RESULT 32

US-08-456-907-4
Sequence 4, Application US/08456907
Patent No. 5633142
GENERAL INFORMATION:
APPLICANT: Herlyn, Meenhard
APPLICANT: Morris, Jennifer
APPLICANT: Rauscher III, Frank J.
APPLICANT: Rodeck, Ulrich
TITLE OF INVENTION: WTI Monoclonal Antibodies and Methods of
TITLE OF INVENTION: Use Therefor
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,907
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/234,783
FILING DATE: 28-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST48AUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-456-907-4

Query Match 4.1%; Score 214.5; DB 1; Length 429;
Best Local Similarity 24.3%; Pred. No. 1.1e-06;
Matches 91; Conservative 43; Mismatches 122; Indels 119; Gaps 18;

133 GGGGGLIASPKLGATP-----LPPEST-----PAPPPPPPPPPGVSGHLNIP 178
19 GGGGGCAL--PVSGAQMVAVLDFAPRGASAYGSLGAPRPPRPPRPHS----- 70
179 LLEELRYLQROQHOMOMTEOICROVLLGSLGQ--TVGA-----PASPSEL-PGT 227
71 -----FIKQPSWGAERHEQCLSAFTVHFGQFTGTAGACRYGPRGPPRPSQASSGQ 124
228 GTASTKPLPLFSPRIKRYQVTSKTLASSSSSSSSSSG-----AETPKQAFHLYHPLG 280
125 ARMFPMAPYLPSCLSEOPAIRNOGYSTVTFDGTSPSYGTHPSHNAQFPHNSFKH-EDPMG 183
281 SQHPFSAGVGSRHKPTP-----APSPALPGSTDLIASPHLA---FPSTTGL----- 325
184 QQ-----GSLGEQYQSVPRPVYGCNTPDSCSTGSQALLKRTPRYSSDNLXQMTSOLECMTW 238
326 ----LAAOCLGAARGLEAT-----ASPGILKPKN 350
239 NQMNLGATLKGHSTGVESDNHTTPIILCGAQRHHTGVRGIDQVRRVGVAPTLVRSAS 298
351 GSGELSYGEVMGLKPRGGRHKCRF--CAKYFGSDSALOIHLSHTGERPYKCNV--CGN 406
299 ETSE-----KRP---FMCAYPGCKNRKYFKLSHLQMSRKHTGKPYQCDKDCER 345

[illegible]

RESULT 37
 US-08-711-417C-197
 Sequence 197, Application US/08711417C
 Patent No. 6228611
 GENERAL INFORMATION:
 APPLICANT: Georgopoulos, Katia A.
 TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
 NUMBER OF SEQUENCES: 202
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows 95
 SOFTWARE: FastSeq for Windows Version 2.0b
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/711,417C
 FILING DATE: 05-Sep-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/238,212
 FILING DATE: 02-MAY-1994
 APPLICATION NUMBER: 08/121,438
 FILING DATE: 14-SEP-1993
 APPLICATION NUMBER: 07/946,233
 FILING DATE: 14-SEP-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Myers, Louis P.
 REGISTRATION NUMBER: 35,965
 REFERENCE/DOCKET NUMBER: 10287/007001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEEX: 200154
 INFORMATION FOR SEQ ID NO: 197:
 SEQUENCE CHARACTERISTICS:

```

; LENGTH: 432 amino acids
; TYPE: amino acid
;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 197
US-08-711-417C-197

```

Query Match	4.0%;	Score 209.5;	DB: 4;	Length 432;
Best Local Similarity	22.8%;	Pred: NO.24e-06;		
Matches	97;	Conservative 43;	Mismatches 136;	Indels 149;
				Gaps 18;

[illegible]

```

RESULT 38
US-09-110-517-2
; Sequence 2, Application US/09110517A
; Patent No. 6248520
; GENERAL INFORMATION:
; APPLICANT: Roeder, Robert G
; APPLICANT: Fondell, Joseph D
; APPLICANT: Yuan, Chao X
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING NUCLEAR HORMONE
; TITLE OF INVENTION: RECEPTOR COACTIVATORS AND USES THEREOF
; FILE REFERENCE: 600-1-224
; CURRENT APPLICATION NUMBER: US/09/110,517A
; CURRENT FILING DATE: 1998-07-06
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1581
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-110-517-2

Query Match          4.0%; Score 209.5; DB 4; Length 1581;
Best Local Similarity 19.3%; Pred. No. 1,3e-05;
Matches 174; Conservative 91; Mismatches 263; Indels 373; Gaps 33;

5 SERSRLGVPAEPALGGDASEEDHPQVCAKCAQFTDPTFLAHONACSTDPVAVII 64
I : : ||||| : ||||| I :| : : : :

```

Db 900 SOALNTLGVPM-----LGDNCE-----TFKGNQADYDFTSIISVA 937
QY 65 G-----GOENP-----NNSASSEPRREGHNP----- 87
Db 938 GKALAPADLMENHSGSOGPLTTGDLGKEKTQKRYKEGNGNSTLSGLDLSKPGKRSR 997
QY 88 -----QVMDHEHNPDSGSSVP--DDPTWGPERRGEESGHFLVAATGTA 131
Db 998 TPSPNDOKSKDKPRKKADTEGKSPSHSSNRPTPTP-----STG-----GSK 1041
QY 132 AGGGGGLILASPKLGTATPLPESTPAPRPPPPPPGSGHLNPLILELRVLQORQ 191
Db 1042 SPGSGN-----RSQTPRGVATPTPT----- 1061
QY 192 IHOMOTEDICROVLLIGSLGQTVGAPASPELPGTASSTKPLPLPFSPIKPVOTSKT 251
Db 1062 ---KITIPIPKGTVM-----VGKPSHGOYTSGSVSSGSS-----KSHNS 1099
QY 252 LASSSSSSSSSGAETPKQAFPHLYHPLGSHQHPFAGGVGRSHKPTPAPSPALPGSTDL 311
Db 1100 HSSSSSSSASTSKMKSSKSSGSSSKLSSMYSGSSGSSOSKSSSGGKPGSS--- 1156
QY 312 IASPHLPSTTGLLAQCLGAARGLEATASPGLIKPKNGSELXYEYWGPR-----LE 365
Db 1157 -----PIT-----KHGLSSGSSSTKMKPGKPPSLMNPISLKPNTSPSHR 1197
QY 366 KKGGRKRCRCACAVFGSDSLQIHLRSHGTGERPYKCNVCGNFTTGNLKVHHRHREKY 425
Db 1198 PPGGSDKL-----ASPMK----- 1210
QY 426 PHVQMPNRPVREHLDYITSGSLPYGMSVPRPEKAEPEATPGGVERKPLVASTTALSAT 485
Db 1211 -----PVP-----GTP-----PSSKAKSPISGSG-----SHMSGT 1237
QY 486 ESLLTSTAGTATAPGLPAFNKFLVLMKAVEPKNADEMTPPGSEGSALISGVAESSTATL 545
Db 1238 SSSSGMKSSGSLGSSGL-----SOKTPPSSNCTASASSPSSSGSS 1279
QY 546 MQLSKMLTSLPSMALLNHFSTGSPFLPLCARALGASPESTKLOLVKEKIDOGAVAV 605
Db 1280 MSSSQ-----NOHGSSKGK-----SPSRNKK-PSLTVAVIDLKRGV 1315
QY 606 TSAASGA-PTTSAPAPASSASSGPRQCVCILRVLSCPRALRLHYGQHGSRPFCKYCGR 664
Db 1316 TSGPGEEDPLDGOGVSTNSSHP-----MSSKHNMSGGE--FGK---R 1355
QY 665 AFSTRGNLRAHFVGHKASPARAONS-----CPICOKKFTNAVTLQ 705
Db 1356 EKSDKDKSVSTSSSVSSKTSSEKNVSTSVAKIITSKHGGSPSIKAK-----VTLQ 1411
QY 706 QHYRMHLGGQPR-----NGSTALPREGGAOENGSEQSVSGA-----GSFP 747
Db 1412 KPGESSGEGLRPMSSKNYSPRLTSGTPKHHERGSPSHSKSPAYTPQONLDESESGSSI 1471
QY 748 QOOSQOQSPPEE-----ELSEEDEEEDDEEDYTDSDSLAGRSESGEKAISVBCD 799
Db 1472 AELSXYQSPSSDDPLRPLREYSTEKHKHKKKKKKKYKDKDRDRDRDKKKHSHIRPE 1531
QY 800 S 800
Db 1532 S 1532

RESULT 39
US-08-711-417C-196
; Sequence 196, Application US/08711417C
; Patent No. 6228611
; GENERAL INFORMATION
; APPLICANT: Georgopoulos, Katia A.
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
; NUMBER OF SEQUENCES: 202
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson P.C.

STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/711,417C
FILING DATE: 05-Sep-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/238,212
FILING DATE: 02-MAY-1994
APPLICATION NUMBER: 08/121,438
FILING DATE: 14-SEP-1993
APPLICATION NUMBER: 07/946,233
FILING DATE: 14-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis P.
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10287/007001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 196:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 196:
US-08-711-417C-196

Query Match 3.9% Score 206.5; DB 4; Length 461;
Best Local Similarity 23.5%; Pred No. 4,1e-06;
Matches 110; Conservative 44; Mismatches 162; Indels 153; Gaps 21;

QY 329 QCLGAARGLEATASPGLIKPKNS-----GELSYGEYWGPLEKPGGRHKRCRCACAVFGSDS 384
Db 22 ECACDLRLMLDASG-----EKMNGSHRDQSSALSGLV-GGIRLPRGKXKDCIGIICGPN 75
QY 385 ALQIHLNHSHTGERPYKCNVCGNFTTRGNLKVHHRHREKYR----- 426
Db 76 VLMVHKRSHTERPRQCNOCGASFQCKGNLRIKILHSGERPFKCHLCNVACRRRDALTG 135
QY 427 ---HYQKNPH-----PVREHL---YVITSGSLPYGMSVPRPEKAEPEATR 466
Db 136 HLRHNSVKKPRHKCGYCGRKYQRTSLSEHKRECHNYLESMDLP-GTLXPVYKEETKHSEM 194
QY 467 GG-----GVERKPL---VASTALATESITLLTSTAGTATAP-GLPAFNKFLVLMKAVEP 517
Db 195 AEDLCKTISERSLVLDRLASNVAKRKSMPQKFLGDKGLSDTPYDSATYERKENEMKSHV 254
QY 518 KNAKADEN-----TPRG-----SEGSAISGVAESSTAT- 544
Db 255 MDOAINNAINYLGAESLRPLVOTPPGGEVYVISPMTQLHRRESEGTFRSHNSAODSAVE 314
QY 545 ---LMQLSKLMTS---LPS-----WALLTNHFKSTGSFPLPLCAR 578
Db 315 YLLLSKAKLVSEKREKASNSQDSTDTSTNNDEQKSGLLYLTLNH-----IARR 364
QY 579 ALGASPESTKLOLVLEKI--DRQGAVAVTAASGADPTSPADSSASSGPNQVCITLR 636
Db 365 AQRVSLKEHRAVLDLRAASENSQDALRVYS-----TSGQMKVYKCEHC-R 410
QY 637 VLSCPRL-RLHYGQHGGERPFCKYCG-----RAFSTRGNLRAH 675
Db 411 VLFIDHVAVTYTHMGCHGFRDPFECNMCGYHSQDRYEFSSHTTRGBRPH 459

```

QY 455 --PEKAESEEAATPG-----GVERKPLVASTALSATSLTLLTSAG--- 496
Db 1074 VCSNPECETHETGTTNATTATATSMAGOHCSNPCEETHETGTTNATTATSMAGANHOR 11333
QY 497 ---TATAPGLPAFNKF-VLMKAVEPNKNKADENTPGSEGSASISGVAASSTALIMQSLKM 552
Db 1134 DARRACAAGTPAIVIRISVATGAL-----AAQSKSQCGTQROTSTSTTMYMA 1182
QY 553 TSLPSMALLNHRKSNGSFPLPCARLAGSPSTSKLQOLVEKIDQGANAVTSAASG- 611
Db 1183 TGAPC-----SKGPLIGPMAREBGRSPAFVOLAPLSSKV-RLSSPSINDLPRG 1232
QY 612 ---APTAPAPSSSSASSGPNOCVIC-----LRVLSCPRALRLHYGONG 652
Db 1233 HSHAVSTAAMTRBSVYAGBPRMAPVCBSLGGSPSTTVYTALEALLCPAT----- 1284
QY 653 GEHPFCKYCGRAFSTRGNLRAHFV---HKASPARAONSC--PICOKKFT---NAV 702
Db 1285 ----VYQVC-----SNPCETHETGTTNATTATSMASAGRYCSNPCETHETGTTHTAT 1334
QY 703 TLQOHVRMLIGLOIPNGGTALPEGG-----GAAQENSGEOST-----VSGASGSPQOQSO 752
Db 1335 TATS-----NGTGGQPGGGQQPARGRCCEHTQTTSTGTMTSVSGALLPDTSS 1383
QY 753 Q-----PSPBEELSEEEBDEEBEDVDSDSLAGRSE 787
Db 1384 HRTVESGLEVMAAPSVTPQAGTALLAPFTQRYCNSNPCEHETGTTHTATVYNSMSSN 1443
QY 788 SGGEKIASVRGBSESEASGAEEVGTAAATAACKENDSNKTTQOSSLPPR---PPDS 843
Db 1444 QDPPRASOGVEESTOGGSVNTTSSALITTVYSSLLTRAVTIVTOSTPVPGGSVPPEE 1503
QY 844 LD-QOPMEMQ-----GSSGVLGKKE---GSKPE-----RSSSPASALPPEG 881
Db 1504 LQVSGCPROOLPPROLLQASATAIMESEAVLSAQTPBELPAVADLSSTGPESSGGSAG 1563
QY 882 EATSVTL-----VEELSLQAMKE-----IGESSKRACEVCGQ 916
Db 1564 SAVVATVIVQPPPOSEVDOLSTPOELMAEAOAGTTTLMVTGLTPEELAVTAAAEAAQ 1623
QY 917 AFPSQAALREHQ 928
Db 1624 A---AATBEAQ 1631

```

Search completed: January 13, 2003, 15:23:14
Job time : 53.0224 secs


```

QY      841  PDSLDOPQPMQSSGVLGKREBKRESSSSPSALTPREGATSVYLVELSLQEMARK 900
Db      841  PDSLDOPQPMQSSGVLGKREBKRESSSSPSALTPREGATSVYLVELSLQEMARK 900
QY      901  EPGESSRKACEVCGQAFPPSQALEENOKTHPEKGBLFTVCFCROGFLERATLKKHMLA 960
Db      901  EPGESSRKACEVCGQAFPPSQALEENOKTHPEKGBLFTVCFCROGFLERATLKKHMLA 960
QY      961  HHQVQPAHPGPNQIALSLIVGCSPTSTGTGSPPRKDDPIF 1005
Db      961  HHQVQPAHPGPNQIALSLIVGCSPTSTGTGSPPRKDDPIF 1005

1
RESULT 4
US-09-988-117-3
; Sequence 3, Application US/09988117
; Patent No. US20020156039A1
GENERAL INFORMATION:
APPLICANT: Benjamin, Thomas L.
APPLICANT: Li, Dawei
APPLICANT: Mok, Samuel C.
APPLICANT: Cramer, Daniel W.
APPLICANT: Ma, Yupo
TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
FILE REFERENCE: 00742/066002
CURRENT APPLICATION NUMBER: US/09/988,117
PRIOR FILING DATE: 2001-11-16
PRIOR FILING DATE: 2001-03-19
PRIOR APPLICATION NUMBER: US 60/216,723
PRIOR FILING DATE: 2000-07-07
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 1002
TYPE: PRT
ORGANISM: Mus musculus
US-09-988-117-3

Query Match      88.6%; Score 4677.5; DB 9; Length 1002;
Best Local Similarity 89.0%; Pred. No. 7e-22;
Matches 894; Conservative 25; Mismatches 83; Indels 3; Gaps 3,

QY      1  MAHSESRSLGVPGEPAELGDSAEEDHPQVCAKCCAAFTPTREFLAHNACSTDPPV 60
Db      1  MAQETGSSSRIGGPEGEREKGDASEEHHPQVCAKCCAAFSDFTEFLAHNACSTDPPV 60
QY      61  MVIIQGENPNNSSSASEPRRPGHNNPOYMDTEHSNPPDSSSVPTDPTWPERGESS 120
Db      61  MVIIQGENPNNSSSASAPRPGHNSQYMDTEHSNPPDSSGSDPPPTWPERGESS 120
QY      121  GHFLVAMGTAAAGGGGILLAPKIGATPLPRESTARRPPRRPPRRPVGSGHNTPLI 180
Db      121  GHFLVAMGTAAAGGGGILLAPKIGATPLPRESTARRPPRRPPRRPVGSGHNTPLI 180
QY      181  LEEELRVLQORIQHOMQTEQIQROYLLSLGQTVGAPASPSLELPGTASSTRPLBLF 240
Db      181  LEEELRVLQORIQHOMQTEQIQROYLLSLGQTVGAPASPSLELPGTASSTRPLBLF 240
QY      241  SPIRVPQVSKLASSSSSSSSGAFTPOAFHVKVPLDSQHFSGVGRSHKPTPAP 300
Db      241  SPIRPAQVQKTTA-SSSSSSSSGAPPRQAFHVLNPLDSQHFSSVGVGRSHKPTPAP 299
QY      301  SPALPGSTDOIILASPHLAFPSTTGLIAAOCLGAARGLAEATASPLKPKNGSGELSYGEV 360
Db      300  SPALPGSTDOIILASPHLAFPSTTGLIAAOCLGAARGLAEAAASPLKPKNGSGELSYGEV 359
QY      361  MGPRLPGGRHKRCRCACVFGSDSAQIHLRSTGEPDYVCNVGNGNFTTGNLKVHHR 420
Db      360  ISSLEKPEGRHKRCRCACVFGSDSAQIHLRSTGEPDYVCNVGNGNFTTGNLKVHHR 419
QY      421  HREKYPRHQMNHPVPEHLDVYITSSGLPYGMSVPEKAEENATPGGVERKPLVASTT 480

```

Db	420	HREKTPHYQOMNHVPPEHNDIVYITSSGLEPYGVSVEPKAEABEAGTIPGGVYEKKPLVASTT	479
Oy	481	AISATSESTLLUSTSAGCTATACGLPAFKNFKVLKMAVEPKNADENTPPGSEGSAIGVAES	540
Db	480	AISATEESTLLSTGSTAVAPGLPFENFNKFLVLMKAEPKSKADENTPPSEGSAIGVADS	539
Oy	541	STATUMLQSLKLTSTLPSMALLTNHRKSGSPFLPLICARALGASPETSKLOOLVKRIDRO	600
Db	540	GSATMQLSKLVTSLPSMALTLTNHLKSTGSPFPFYVLBJPGASPSBTSKLOOLVEKIDRO	599
Oy	601	GAVANTSAASAGPTTSARPSSASSGPNOCYTCLRYLSCPALNLJHGQHGEERPFCK	660
Db	600	GAVAASTASGAPTTSARPSSA-SGPNOVCYLRLVSCPALRKHNGQHGERPFCK	658
Oy	661	VCGRAFSTRGNIIRAFVFVGHKASPARAONSCPIOCKKEFNATLLOOHVMHJGGIIPNG	720
Db	659	VGGRAFSTRTGNLRHFVGHKTSPARAONSCPICOKKFNTAVTLLOOHVMHJGGIIPNG	718
Oy	721	TALPRGGAAONGSEOSTVSGASFPOOOSOPBEELSEEEDDEEDVDIDS	780
Db	719	SALSREGGAOENSSEQSTASGPSPOSOQPSEEMWS-EEBEDEDEEDVDIDS	777
Oy	781	IAGRSESSEGEALISVRGDSEEAAGAEEVGVAAATAGKEMDSNEKTIQQSSLPPRP	840
Db	778	LARGSESSEGEKALISVRGDSEEVSGAEEVATSVAAPTYVKEMDSNEKAPDHTLPPRP	837
Oy	841	PSTLDPOPMEGSSGYLGKKEGGKPRESSSPAALTPEGATSVTLVEELSLDEAMRX	900
Db	838	PNLDPPOPMEGTISVSGAMEEAKLBETISSPMALTOEGGSTPLVEELINLPBAKK	897
Oy	901	EDGESSRRKACEVCGQAFPSQALEEHOKTHPKEGDLFTVCRCOGFLERATLKRMILA	960
Db	898	DGESSEGRKACEVCGQSFEPTOTALHEHOHTHKDKGLFCVCFROGFLDRATLKRMILA	957
Oy	961	HHOVOPFAHPGPNITAALSLVPGCSPTSITSLSPRRDDPTIP	1005
Db	958	HHOVPPFAHPGPNITATLSLVPCOSSITSPQLSPPRRDDPTMP	1002
RESULT 5			
US-09-812-471-3			
; Sequence 3, Application US/09812471			
; Patent No. US20020018765A1			
; GENERAL INFORMATION:			
; APPLICANT: Benjamin, Thomas L.			
; TITLE OF INVENTION: Diagnosing and Treating Cancer Cells			
; FILE REFERENCE: 00742/062002			
; CURRENT APPLICATION NUMBER: US/09/812,471			
; CURRENT FILING DATE: 2001-03-19			
; PRIOR APPLICATION NUMBER: US 60/7216,723			
; PRIOR FILING DATE: 2000-07-07			
; NUMBER OF SEQ ID NOS: 21			
; SOFTWARE: fastseq for Windows Version 4.0			
; SEQ ID NO 3			
; TYPE: PRT			
; ORGANISM: Mus musculus			
US-09-812-471-3			
Query Match 88.6%; Score 4677.5; DB 10; Length 1002;			
Best Local Similarity 89.0%; Pred. No. 7e-222;			
Matches 894; Conservative 25; Mismatches 83; Indels 3; Gaps 3.			
Oy	1	MAHERSRRLVPAAGEPALGLGDASEEDHPVCAKCCAQFTDPTFEFLAHONAGSTDPPV	60
Db	1	MAOETGSSRLGAPCGEPARERGDASEBHNPVCAKCCAQFDPTFEFLAHONSCTDPPV	60
Oy	61	MVIIGQENPNNSASSSEPREGHNHPVOMTEHNSPPRSSGVLPDPTWCPERGEES	120
Db	61	MVIIGQENPNNSASSAPRPGHSKVQVMDIEHSNPDPDSSGGSPDPTWCPERGEES	120

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 43157
LENGTH: 336
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL034420.16
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.72
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.75
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.76
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.78
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.75
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.86
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.81
OTHER INFORMATION: EST_HUMAN HIT: AM270757.1, EVALUE 7.00e-68
OTHER INFORMATION: SWISSPROT HIT: P39806, EVALUE 8.00e-39
US-09-864-761-43157

Query Match 8.4%; Score 441; DB 10; Length 336;
Best Local Similarity 34.3%; Pred. No. 3.4e-15;
Matches 119; Conservative 41; Mismatches 89; Indels 98; Gaps 11;

585 SETSKLOQLEKIDROGAVAVTSAASGAPTTSAAPSSASSGPNOCYICLRVLSCPRAL 644
62 SETKLQQLVENDK-----ATTDPNECLDCHRVLSGSSSL 97
645 RLHYGCHGGERPFKCKYCGRAFTNGNLRAHFVGHKASPARAARNSCPICQKKTNAVTL 704
98 KMHYRTHTGERPFCKKRASTKGNLTKHVLGVRHTNTSITKTHSCPIQCKKFTNAVNL 157
705 QQHYRMHLGGQIPNGGTALPEQSGAAQENGSEQTVSAGSFPOQSOQSPSEELSEEE 764
158 QQHIRMHGGQIPN--TPPE--NPDFTGSEPMFTVGENGS-----TGAICHD 202
765 EEDDEEEDVYDDEDSLAGRSEGEKKAISVRGDSSEASGAEEVGV-----AAATA 819
203 VIESIDVEE-----VSSQEAPESSSKVPTPLPSIHSASPTL 238
820 GKEMDSNKTTOQSSLRPP-----PPDSLDOPQPMEOG-----SSGVLGGESEEG 865
239 GFAM-----MASIDAPGKKGPAPFNLDORGRSRENGSVSDGILTNTSSSLMGDOE--- 287
866 KPERSSSPASALTPEGEATSVTLVEELSLQEAEMRKREGESSSRKACE 912
288 --YQSRSPDILETTSFGALSPANSQAESINS--KSPDAGSKAESSE 329

RESULT 9
US-09-864-761-37152

Sequence 37152, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 37152
LENGTH: 515
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL049646.17
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.99
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EST_HUMAN HIT: AW953464.1, EVALUE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: P52736, EVALUE 0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: AW953464.1, EVALUE 0.00e+00
US-09-864-761-37152

Query Match 5.1%; Score 267.5; DB 10; Length 515;
Best Local Similarity 20.9%; Pred. No. 1.6e-06;

Matches 160; Conservative 60; Mismatches 232; Indels 313; Gaps 30;

OY 237 LPLPSPPIKPVQTSKLTIASSSSSSSSSSSGAETPKQAFPHLYHPGSGHPFSAAGVGRSHKP 296
Db 13 MFLGRTK---KRTLGAFS-----RPPQR---QPVSSRN---GLGVELEA 49

OY 297 TPAPSPALPGSTDLI-----ASPHLAPPTGLLAACGLGARGLEATASP 343
Db 50 SPFAQS-GNPESTDILKRIEVLGFVTNCGEGSLSTSKMTNLSHQ----- 94

OY 344 GLLKPKNGSGELSTGEYWG-----PLEKPGGRHK-----CRFCAKVFSGDSALQ 387
Db 95 -----RIHSGEKPY--VCGVCEKGFSLKSLARHOKAHSGEPYVCEGGRGFNRKSTLI 147

OY 388 IHLRSHGERPRYKCNVCGNRFRTTGNLKVHFRHREKYPHYQNNPHVPRHLDVYITSSG 447
Db 148 IHERTHSGEPRYMCSECGRGFSQKSNLTIHORTHSGEKPY-----YVCRECG 193

OY 448 LPYGMSVPRPEKAEBEAATPGGVERKPLVASTATLATESLTLLSTAGATAPGLPAFN 507
Db 194 KGFSGKSAVVRHQRT-----HLEKTIYVSDCGLGFSDRSNLIS----- 232

OY 508 KPYLMAVEKKNADENTPPSGEGSAIS--GYAESSTATLMQSLKMTSLPWSALLTNHF 565
Db 233 -----HORTHSGEKPYACKECGRCFRQRTTLV-----NH- 261

OY 566 KSTGSPFLPLCARALGASPSSETSKLOQLVEKIDROGAVAVTSAGAPTTAPAPSSSAS 625
Db 262 QHTHSKEKPYVCGVCGHSFQNSTL-----ISHR-----RTHTG 295

OY 626 SGPNOVCYICLVLSCPRALRLHYGOHGERPKCKVCGRAFSTRGNLRAHFVGHKASPA 685
Db 296 EKPYVCGVCGRGFSQKSNLVAHORTHSG-----ERYVCEG-- 351

OY 686 RAONSCPLOCKFTFNNAVTLQOHVBMHGGQIPNGSTALPEGGGAADENGSEOSTVSAGS 745
Db 352 EKPYVCGVCGRGFSQKSNLVAHORTHSG-----ERYVCEG-- 389

OY 746 FPQOOSQOPSPEBELSEEEDEEDEDVTDSDSLAGSGESGEEKAISVRGDSSEASG 805
Db 390 -----RQFSHQ----- 395

OY 806 ABEVGTVAATAAGKEMDSNEKTTQOSSLP PPPPPSLDOPQPMEGSSGVLGKEEG 865
Db 396 -----AGLIRHKRKHREK-----PYMCRQCGLG-FGNK----- 423

OY 866 KPERSSPASALTPPEGATVTLVEELSLQAMRKPEGSSSRKACEVCGQAFPSQALE 925
Db 424 -----SALITHKRAHS-----BEKP-----CYCRECGQGLQKSHLT 455

OY 926 EHOKTHPKEGPLFTCVFCRCROGFLERATLKKH--MLAHNQ--VOP 966
Db 456 LHOVMHTGEPK--YVCKTCGRGFSQKSNLHSHRKTTSVHHRLPVOP 499

RESULT 10
US-09-989-920-212
Sequence 212, Application US/09989920
Patent No. US20020172957A1
GENERAL INFORMATION:
APPLICANT: Maclean, Roberto
APPLICANT: Recipon, Hervé
APPLICANT: Chen, Sei-Yu
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
FILE REFERENCE: DEX-0291
CURRENT APPLICATION NUMBER: US/09/989,920
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/252,500
PRIOR FILING DATE: 2000-11-22
NUMBER OF SEQ ID NOS: 284
SOFTWARE: PatentIn version 3.1

; SEQ ID NO 212
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-989-920-212

Query Match 5.1%; Score 267.5; DB 9; Length 610;
Best Local Similarity 20.4%; Pred. No. 1.9e-06;
Matches 129; Conservative 74; Mismatches 187; Indels 241; Gaps 23;

OY 365 EKPGRHKRCFCACVYFGSDSALQIHLRSHTGERPRYKCNVCGNRFRTTGNLKVHFRHREK 424
Db 132 ERP--YVCDVCGFTFRNAGLKVHRLHTGKERPYCDVCGKAYISNLSLKNHGTHLGE 188

OY 425 YP-----HVQNNPHVPRHLDVYITSSGLPYGMSV----PPEK 458
Db 189 KPYKCYCEKSFNYSALQEHKRIHTREKPCDECGKAFRNNSGLKVHRIHTGERPYK 246

OY 459 AEEEAATPGGVEKRPPLVASTATLATESLTLLSTAGATAPGLPAFNKAVLMAVEPK 518
Db 249 CEE-----CGKAYISLSLHNKSVHGERPF----- 275

OY 519 NKADENTPPSGEGSAISGVAESSTATLMQSLKMTSLPWSALLTNHF--STGSPFLPL 575
Db 276 -KCDE-----CEKAFIT-----YRITLHNKRVHLEKPYKCDV 307

OY 576 CARALGASPSSETSKLOQLVEKIDROGAVAVTSAGAPTTAPAPSSSASGPNQCYICL 635
Db 308 CEKSF--NVTSLLSOHRVHTRE-----KPYECDRCE 337

OY 636 RVLSCPRLRLHYGOHGERPKCKVCGRAFSTRGNLRAHFVGHKASPARAONSCPICQ 695
Db 338 KVRNNSSLVAKHRIHTGERPYECDVCGKATISHSL-----INHKSTHPEKTHTCDECG 393

OY 696 KFTNAVTLQOHVBMHGGQIPNGSTALPEGGGAADENGSEOSTVSAGSFPQOOSQOPS 755
Db 394 KAFSSRTLLSHKRVHIG-----EKPFKVCVEGKSFSSYS-- 428

OY 756 PEEBELSEEEDEEDEDVTDSDSLAGSGESGEEKAISVAGDSSEASGAEEVGTVA 815
Db 429 ----LSQHKRIHTGERPYVCDRCGKAFFNS-----S 456

OY 816 AATAGKEMDSNEKTTQOSSLP PPPPPSLDOPQPMEGSSGVLGKEEGKPERSSPAS 875
Db 457 GLTVHKKRIHTGK-----PYE-----CDECGKATISHS--- 484

OY 876 ALTPREGATSVTLVEELSLQAMRKPEGSSSRKACEVCGQAFPSQALEEHOKTHPKEG 935
Db 485 -----SLINHSVHOG--KOPYN-----CE-CGKSFNYSVLDQHKRIHTGKK 524

OY 936 PLFTCVFCRCROGFLERATLKKHMLAHNQVOP 966
Db 525 P-YRCNDC-AHLPNAT--ADLMKVVDHEEP 550

RESULT 11
US-09-764-864-883
Sequence 883, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 883
LENGTH: 577
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-864-883


```
; Sequence 948, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 948
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (335)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-948
```

```
Query Match          4.6%; Score 244.5; DB 10; Length 338;
Best Local Similarity 23.8%; Pred. No. 1.4e-05;
Matches 86; Conservative 34; Mismatches 145; Indels 97; Gaps 10;
```

```
OY 366 KPGRRHKRCRCARVFGSDSALOIHLSHTGERPYKCNVCGNFTTGTGNLKVHFHREK 425
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 40 KP---HSCPRCSKLFKQPSHLQTHLTHQGRPHKQYCHKAFTQSLKRNHMLHSEVK 96
OY 426 PNYQMPNHPYREHLDVYITSSGLPYGMSVPRPEKAEAEATPGG-----GYERKPLVA 477
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 97 PY-----SCHFCGRGFAYPSELKAHEVKNHESGRCHVCVEGDFSTLTQ 140
OY 478 STTALSTESLTITLSTAGTATAPGLPAFNKFVLMKAVEPKNAKADENTDP-----GSEG 531
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 141 LKRHLASHQSPITLXQC-----LECDKSFHYTSQLONHMLKHONVAPFVCTECMEF 191
OY 532 SAISGVAESSTATLMOLSKLMTSLPSWALLTNHFKSTGSPFLPLCARLGASPSSTKLO 591
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 192 SOIHLKHQS-----LTH--KGVEKFECEVCGREF--TLQANMKRH 228
OY 592 OLVEKIDRGAVAVTSAASGAPITSAAPASSASGPNOCVLCRLVSCPRALRLHGOH 651
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 229 MLIH-----TSVRYQCHICFKEFYVQKQTLKTHMIVH 260
OY 652 GGERPEKRCVCGRAFSTRGNLRAHFVGHKASPARAONSCPICOKKFTNAVTLOOHVRMH 711
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 261 SPVKPFCVKVCGKSFNNMYLGLMHMLHAGSKPRF-----CPYCSSKFNLKGNLSRHMKVK 711
OY 712 LG 713
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 317 HG 318
```

```
RESULT 14
US-09-764-864-976
; Sequence 976, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 976
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-976
```

```
Query Match          4.6%; Score 244; DB 10; Length 406;
Best Local Similarity 17.8%; Pred. No. 1.8e-05;
Matches 106; Conservative 48; Mismatches 145; Indels 296; Gaps 15;
```

```
OY 365 EKPGRHKRCRCARVFGSDSALOIHLSHTGERPYKCNVCGNFTTGTGNLKVHFHREK 424
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 98 EKP---YECPECGAFSEKSRKRRKHQRTHTGKRPYKCDGDKAFSAKSGRLRIHQRTHGE 154
OY 425 YPNYQMPNHPYREHLDVYITSSGLPYGMSVPRPEKAEAEATPGGVERKPLVASTALSA 484
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 155 KP---FECHCECKSFENY-----FPCNECKSPSHSG----- 168
OY 485 TESTLSTSAGTATAPGLPAFNKFVLMKAVEPKNAKADENTPPGSBSAISGVAESSTAT 544
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 169 -KSLIIVHQRTHGEKP-----FPCNECKSPSHSG----- 199
OY 545 LMOLSKLMTSLPSWALLTNHFKS--TGSEPLPL--CARLGASPSSTKLOOLVEKIDROG 601
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 200 -----LENHRRTHTGERPYKCDCEGKAF-----KLKSGLRKHNR-- 233
OY 602 AVAVTSAASGAPITSAAPASSASGPNOCVLCRLVSCPRALRLHYGHGGERPFCKV 661
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 234 -----THTEKPYKCNOCGRKFGKQSOLRGHHRHTHTEKPYKCNH 273
OY 662 CGRAFSTRGNLRAHFVGHKASPARAONSCPICOKKFTNAVTLOOHVRMLGGQIPNGGT 721
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 274 CGEAFSGKSNLRVYHNRH-----TGKPYOCCECGKTFRKSNIIRGQORTH----- 319
OY 722 ALPEGGGAOENGSEOSTVSGAGSFPOOQSOPSPREELSEEEDEDEEDVTDEDL 781
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 320 ----- 319
OY 782 AGRGSESGGKALSVRODSEASGAEEVGTVAANAATACKEMDSNKTQOSSLPPPPP 841
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 320 -----TGKPYECNCC----- 330
OY 842 DSLDOPQMEQSSGVLGKEEGKPERSSPASALTPGCAATSVLVEELSLQEAARKE 901
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 331 -----GKNFS-----EKSVLKRRH 347
OY 902 PGSSSSRKACEVCGQAPPSQALAEHOKTHPEKGLFTCVFPRGGLERATLKKH 956
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 348 TGEKPYN--CNOCGEAFSOKSNLRVHQRTHGEKP--YKCDKCGRFPSSOKSLREH 399
```

```
RESULT 15
US-09-864-761-33476
; Sequence 33476, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
```


PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 33476
LENGTH: 439
TYPE: prt
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AF118808.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.7
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.4
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3
OTHER INFORMATION: SWISSPROT HIT: P52742, EVALU0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: BE885790.1, EVALU0.100e-116
OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALU0.100e-115
US-09-864-761-33476

Query Match 4.68; Score 242; DB 10; Length 439;
Best Local Similarity 22.68; Pred. No. 2.4e-05;
Matches 104; Conservative 43; Mismatches 209; Indels 104; Gaps 13;

294 HKPRPAPSPALPGSDIOLIASPHLAPSTTG---LLAACCLGAAGLEATASPGLKPK 349
53 HREYQVLSQSNP-----LPHQAVPSEGRPYMCYECGCGFR----- 89
350 NGSGELSYGEVYMRPLEKPGRRKRCFCAVFGSDALQIHLRSHTEGPRYKCNVGNRFT 409
90 --SSHLLOHQRHITGEKPR---YVCSVCGKAFSSSVLSKRRIRHITGEKPYCEGKAFR 144
410 TRGLKLVHFNHREYVYQNMHPHYREHLDYITSSGLYGYKMSVPREKAELEATPBG 469
145 VSSLOAHNKHITGKRPHECLECRKAFQLSLIQHQRHITG----- 186
470 VERPLVASTALSTESLTLLSTAGTAPGLPAFNKFVLMKAVEPRKNKADENTPPGS 529
187 --ERPYCPRLCGKAAHNSHTVLRSR-----VHTGEKPHICNE----- 222
530 EGSAISGVAESSTATLMOLSKMTSLPSW-----ALLTNHFKSTGSPPLPLC 576
223 ----CGKTFVKKRTLQHRHITGENTPCTSECGKAFSDRSYVLIQHNHVTGKPYE-C 276
577 ARAUGASSESESKLOQVLEKIRGCAVAVTSAASGAPTTGAPARASSASSG--PQCVCIC 634
277 SEC-GKTFSSHSTLMNHERIHTEEKPYACEGKAFYOVSHLLOHQRVHTGKRPYVCGEC 335

635 LRVLSCPRALRLAHYGOHGEPRPFKVCGRAFSTGRLRAHFVGHKASPARAONSCPIC 694
336 GHAFSARSLIQRHRIHTEKPRFQCTCEGKAFSKRTLIVLHRTHT---TGKPYECNSC 391
695 QKFTNATVTLQOHVRAHMLGGOLPNCGTALPEGGAQENG 734
392 GKAFSQYSVLQHRHITGEKPYECG---ECGRAFNHG 427

US-09-864-761-42897
Sequence 42897, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Acomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 42897
LENGTH: 492
TYPE: prt
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007228.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
OTHER INFORMATION: SWISSPROT HIT: P10078, EVALU0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: AL042491.2, EVALU0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: AL042491.2, EVALU0.00e+00
US-09-864-761-42897

```

? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: PCT/US01/00666
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00667
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00664
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00669
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00665
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00668
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00663
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00662
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00661
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00670
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: US 60/234,687
? PRIOR FILING DATE: 2000-09-21
? PRIOR APPLICATION NUMBER: US 09/608,408
? PRIOR FILING DATE: 2000-06-30
? PRIOR APPLICATION NUMBER: US 09/774,203
? PRIOR FILING DATE: 2001-01-29
? NUMBER OF SEQ ID NOS: 49117
? SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
? SEQ ID NO 36336
? LENGTH: 375
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AC011451.2
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
? OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
? OTHER INFORMATION: EXPRESSED IN B1474, SIGNAL = 1.8
? OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
? OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5
? OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7
? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.96
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
? OTHER INFORMATION: EST HUMAN HIT: AF150239.1, EVALUE 1.00e-102
? OTHER INFORMATION: SWISSPROT HIT: Q14585, EVALUE 1.00e-124
US-09-864-761-36336
Query Match 4.58; Score 237; DB 10; Length 375;
Best Local Similarity 24.14; Pred. No. 3,6e-05;
Matches 84; Conservative 43; Mismatches 170; Indels 52; Gaps 8;
OY 372 KCQFCAKVQESDSALQHLHRSHTGTERPKVCNVCGRNFTTGNLKVPHNRHREKRYHQMN 431
Db 39 ECECKCAFAFDHSSLKSHIRSHTSKPYQCKECGKAFAHFLACPKKHNKPTRE-----K 92
OY 432 PHPVPEHLIDVITSSGLPYGMSVPEKAEAEATPGGVEERKPLVASTALSAATESITIL 491
Db 93 PEYCKRCOTKAFSCSSSFPRAHMKIHIGKTNYECKEGKG-----FSCSSSILT-- 138
OY 492 STSAGTATAGLPAFNKFTVLMKAVEPKKADENTPPGSEGSALISVAE---SSTATIMQ 547
Db 139 -----EHKRIHSGDKPEYCKECGKAFCSSSSLSLKNRISHGDKPYE 179
OY 548 LSKLMTSLPSWALLTNHFK--STGSFPLDLCARLGAASPESTKL--QDLVEKIDRGAVA 604
Db 180 CKECGAFASSSHLIIRIHTGKPYE--CKECKAISESSSLVYHGRTHTGKPYCK 237
OY 605 VTSAAAGAPTTAPADSSASGPNOCYICLDVLSCPRALRIHYGONGBERPRPKVCYGR 664
Db 238 ECGKAVNCSSSLITHMKHKGKPYECUECGKAFLPTSLNTHVKMQSREKPEYCKECGK 297

```

Oy 665 AESTGNLNAHFVGHKASPAARQNSCPICQKFTNTAVTLQOHVPMHLG 713
Db 298 AFSCHSSFRAHVRDH---TGKIYCECKGCKTFSRSSSLTEHLRTHSG 342

RESULT 18
US-09-864-761-34333
Sequence 34333, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Penn, Sharon G.
APPLICANT: Hanzel, David R.
APPLICANT: Hanzel, David R.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-x-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 34333
LENGTH: 311
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC008165.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.97
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7

OTHER INFORMATION: SWISSPROT HIT: P52742, EVALU 1.00e-115
OTHER INFORMATION: EST_HUMAN HIT: A0133067.1, EVALU 3.00e-85
US-09-864-761-34333
Query Match 4.4%; Score 234.5; DB 10; Length 311;
Best Local Similarity 23.5%; Pred. No. 3.9e-05;
Matches 85; Conservative 37; Mismatches 142; Indels 97; Gaps 11;
Oy 365 EKPGGRHKCFKCAKVFSGDSALQIHLRSHTGPRPKCNVCNFTTNGNLKVHHRH-RE 423
Db 27 EKP---HGCHLCGKAFTHCSDLRKHERHTHGEKPYGCHLCGKAFSSNLRRHEMHTRE 83
Oy 424 K-----YPHVQNMHPHVEHLDVYITSSGLPYGNSVPEKAEAEATPGGVK 473
Db 84 KAQICHLGKAFTHCSDLRKHERHTL-----GDKPYG----- 115
Oy 474 PLVASTALSTESITLTLSTAGTATDGLPAPNKFVLKMAVEPKKADENTPPGSGSA 533
Db 116 -----CLLCGK-----AFSKCSYLROHERHTHNGEKPVECHLCGKA 150
Oy 534 ISGVAESSTATLMQSLKMTSLPSMALLTNHPS--TGSFP--LPLCARLGAASPESTKL 590
Db 151 FSHCSH-----LROHERSHNGEKPVECHLCGKA--TESSVLKR 187
Oy 591 QQLVEKIDROGAVAVTSAASGAPTSAPAPSSASSGPNQVCITLRYLSCPRLRLHYGQ 650
Db 188 HERHTHGEKPYECHGKAFTHCSDLRKHERHTHGEKPYGCHLCGKAFNHSVLRHERBT 247
Oy 651 HGERPFRCKVCGRAFSTRGNLRHAFVGHKASPAARQNSCPICQKFTNTAVTLQOHVPM 710
Db 248 HTGEKPYECNICGKAFNRSYNFRLHRRVH---TGKPYVCPILCGKAFSKFENLRHERBT 303
Oy 711 H 711
Db 304 H 304

RESULT 19
US-09-864-761-35894
Sequence 35894, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Penn, Sharon G.
APPLICANT: Hanzel, David R.
APPLICANT: Hanzel, David R.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecm1ca-x-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35894
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF166490.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUATE 1.00e-91
; OTHER INFORMATION: SWISSPROT HIT: P52742, EVALUATE 1.00e-118
US-09-864-761-35894

Query Match 4.4% Score 234.5; DB 10; Length 340;
Best Local Similarity 23.5% Pred. No. 4.3e-05;
Matches 85; Conservative 37; Mismatches 142; Indels 97; Gaps 11;

365 EKPGRHKRCFCAKVFGSDSALQIHLRSHTGERPYKCNCVGRFTTNGNLKVFHNH-RE 423
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
56 EKP---HGHLGCKAFTHCSDLRKHERHTGCKRYGCHLCGKAFKSSMLRRHEMTHTRE 112
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 K-----YHVQMNPHRPREHLDYVITSSGLPYGMSVPERKAEBAATPGGVERK 473
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
113 KAQICHGKAFHCSDLRKHERHTL-----GDKPYG----- 144
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
474 PLVASTLATESLTLLSTASGATAPGLPAFNKFLVLMKAVERPKNKADENTPPGSEGS 533
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 -----CLICGK-----AFSKCYLROHERHTNGEKPVECHLCGKA 179
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
534 ISGVAESSTATTMLQSLMTSLPSWALLTNHFKS-TGSPF--LPLCARALGASPSETSKL 590
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
180 FSHCHS-----LRQERSHNGEKPRGCHLCGKAF--TESSVLKR 216
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
591 QQLVAKIDROGAVAVTSASGARTTSAPADSSSASGPNQCVICLRVLSCPRALRLHYQO 650
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 HERHTGKPYECHVCGKAFTESSDLRRHERHTHGEKPYECHLCGKAFNHSSVLRHERHT 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
651 HGERPFKCVCGRAFTSTGNLRAHFVGHKASPAARQNSPICOOKKTTNATLQOHVEM 710
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 HTGEKPYECNICGKAFNRSYNFRLLHRVH-----TGEKPYVCPGCGKAFSPFLNRQHER 332
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
711 H 711
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
333 H 333
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

; GENERAL INFORMATION:
; APPLICANT: Chen, Huei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 114
; LENGTH: 1125
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 3267030CD1
US-09-974-298-114

Query Match 4.4% Score 233.5; DB 9; Length 1125;
Best Local Similarity 20.3% Pred. No. 0.00017;
Matches 208; Conservative 110; Mismatches 389; Indels 319; Gaps 44;

2 AHESENSSRLGVPAQ-----EPALGSDASEEDH-----POYCAKCCAOFTDP 44
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
222 SHKSGRDQRRHTVQSGCNKRFKTECGKAFKYKHNLKHLRHSNGEKPYECNCKKRSHS 281
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
45 TELPLAH--QNACTSDPYVMVIG-----GOENPNSSASSEPREGHNNPOVMDTEHSNP 98
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
282 GSYSSHSSKRCISLIPVN---GRPRTGLKTSQCSSPSLSASGSPTRPQIROKIEKKPL 338
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99 DSGSSV-----PDDPMGPRRGEBSGSHFLVATGTAAQ-----GGGGLILASP 143
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
339 QEDLSVNOITPEVDYEFK-----IVASGICNCPILONGVTTGGPPLQATSS 387
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
144 KUG---ATPLRP-----ESTPAPPPEPPPP 166
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 PQGMQAVVLPVTVGLVSPISINLSDIQNLKAVDGNVIRQVLENNQANLASKEQETINA 447
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 PPGVSGH-----LNIPLI-----LE---ELRVLOQRIHQOMTEQICROY 205
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
448 SPLOQGGHSYSAISLPLVDDGTTKIITINYSLEEPSQLOVVQNLKKNPVAATNSCKSE 507
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
206 LLLGSLGQTVGAPASPS-----ELPQTGTA-----SSTRPLPLF 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
508 KLPEPL--TVKSEKDKSEFEGGVNDSTCLLDCDCEPDINALPRLKHYDLKQPTPPPLPA 565
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 SPFKPVQTSKTLASSSSSSSSSGAETPKQ-----AFHILYHPLGSO----- 282
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
566 EAERP-----ESSVSATGDNLSFQPLKMLNLKAYVALMAOPSAEELSKIAD 617
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 -----HPFSAGGVG--RSNKP-----TPAPSPALPGSTND----- 309
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
618 SVNLPLDVYKKMEKQAGQISVOSSSEPPRGKVNTPAKNNDOPOSANANEPDSTYN 677
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
310 -----QLIASPHLAFSTTGLLAQCLGAARGLBATASP--GLLKPKNSEGLSYGEVMGP 363
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
678 IQSPDKMTNSPVLPGVST-----NGSRSSTPSPPLNLSSRRMTGYLYTAE--GA 727
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 LEKPGGRHKRCFCAKVFGSDSALQIHLRSHTG---ERYKCVCCNRR--TTRGNLKYHF 418
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
728 QEEP-----QVEPDLSLPKOQGLLERSTISVYQNSVYSVOEELPLNSC 773
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
419 HHRREKYPHVQMNPHRPREHLDYVITSSGLPYGMSVPERKAEBAATPGGVERPLVAS 478
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
774 AKKEPQKDCVTDSEVVN-----VTPSANPINITAIPVTYQOLPTIV--AIADQNSVPC 826
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
479 TTALATESLTLLSTASGATAPGLPAFNKFLVLMKAVERPKNKADENTPPGSEGSASIGVA 538
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
827 LRALANKQTIILIPQAVAYTYSTVSPAVQEPPL-KVIOQPNQNOBERQDTSSSG--VSANE 883
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
539 ESSTATMLQSLMTSLPSWALLTNHFKSGSPFLPLCARALGASPSETSKLOQLVERKD 598
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 884 DQNDSDTPPKKKMRK-----TENGYACDLCDKIFQKSS-----LLRHYE 926
Qy 599 RQNAVAVTSAASGAPTSAAPASSASSGPNOCVLCRLVLSCRALRLHYGHGGERPFK 658
Db 927 HTG-----KRPHECGICKKAFKHKHHLLEHMRHLSGGERYQ 962
Qy 659 CKYCGAFSTRGNLRAHFVGHKASPAAR-AQNSCPICOKFTNAVTLQOHVRLHGGQIP 717
Db 963 CDKCGKRFSGHSISQH-MNHRYSYCKREAREDRSTQOEAGEPILSNENH-----1012
Qy 718 NGGTALPEGGAOENGSEOSTVSGAGSFPOOQSOQPSPEELSEEE-----EDEE 770
Db 1013 -GARASPSQGSQSDR---ESTLREDEDESEKEEEDKEMEELQEKKECPQGDDEEE 1068
Qy 771 EEDVDYDEDSLAGSGSESGEKRAIS---VRGDSF--ASGAEEFVGVAAATAGKENDS 825
Db 1069 EEEVEVEEEV---BEAENEGEAKTEGLMKDDRAESQASSLGQKVG-----ESSEQVS 1118
Qy 826 NEKTTQ 831
Db 1119 EEKTFNE 1124

RESULT 21
US-09-864-761-37065
; Sequence 37065, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37065
; LENGTH: 551
; TYPE: prt
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005261.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: SWISSPROT HIT: P51522, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUE 1.00e-105
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUE 1.00e-105
US-09-864-761-37065

Query Match 4.4%; Score 233; DB 10; Length 551;
Best Local Similarity 21.7%; Pred. No. 8.4e-05;
Matches 97; Conservative 46; Mismatches 193; Indels 110; Gaps 12;

Qy 365 EKPGRNKRCFCAGKAVGSGSALOILHRSHTGERPYKNCVGNRRTTGNLKVHFRH---421
Db 180 ERP---YECMECGKAFNRKSYLTQHORHSGEKPYKCNCGKAPTRHSNFVLIHNRHNGE 236
Qy 422 -----REKYPHYOMNPH-----PVEPHLDYVI-----TSSGLPYGM 452
Db 237 KSFVCTEGCGVFRHRPGLRLHYVHSGENPYECLECKVFKHRBYLIHMOQTHTG-----291
Qy 453 SVPRKAEAEATPGGVERKPLVASTLSTATESLTLSTAGTAPARGLPAFKKFLVM 512
Db 292 -----EKPYECSECGKVFLESALIH---HYVHTGKPFCELECG 329
Qy 513 KAVEPKN--KADENTPPGSEGSALSGVAESSTATLMLDSKMTSLPSWALLTNHKSSTS 570
Db 330 KAFNHRSYLKRQRIHTGKRPVCS-----ECGKAFYTHCSTYLHKKRAHTGEK 378
Qy 571 PPLPCARALGASPESTSKLOOLEKIDROGAVAVTSAASGAPTTSAAPSSASSGPNQ 630
Db 379 FECKRCGKAF-----SNRKDLIRH-----SIHTGEKPYE 408
Qy 631 CVICLRVLSCPRALRLHYGHGGERPFKCVCGRAFSTRGNLRAHFVGHKASPAARAQNS 690
Db 409 CVECGKAFTRMGSLTRHKKRIHSGEKPYECVCEGKSCFCWSTMLIRHAIHTGKPYK----464
Qy 691 CPICOKFTNNAVTLQOHVHMHLLGGOIPNGSTALPEGGGAOENGSEOSTVSGAGSPPOO 750
Db 465 CSECGKAFSRSSSLVQHOHRHT-GKNPISVTDVGRPPTSGOTSVTLRELLGKDFLNVTT 523
Qy 751 SQQPSPEELSEEEDEDEEEDVT 776
Db 524 EANILPEITSSASDQPYQRETPQVS 549

RESULT 22
US-09-864-761-47724
; Sequence 47724, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761

```
/ CURRENT FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 47724
/ LENGTH: 558
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC011451.2
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
/ OTHER INFORMATION: SWISSPROT HIT: P51523, EVALU0 0.00e+00
/ OTHER INFORMATION: EST HUMAN HIT: AF150239.1, EVALU0 6.00e-92
/ OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALU0 6.00e-92
US-09-864-761-47724

Query Match 4.4%; Score 230; DB 10; Length 558;
Best Local Similarity 19.0%; Pred. No. 0.00012;
Matches 118; Conservative 65; Mismatches 204; Indels 234; Gaps 21;
```

```
Db 333 -----THAGEKPYECMCKGKRAFT 350
QY 640 CPRALRLHYGCHGERPRPKCYCGRAFSTRGNIRAFVGHKASBPAAQNSCPICQCKFT 699
Db 351 ERSYLTKHLRRHSGERPKYCKCKGKRAFTEERSDLTKHLRRHTGDKPYEKD----GKAFV 406
QY 700 NAVTLQOQVHRMHLGGQINGTALPEGGAQOENGSEQSVSAGSPFQQOQSPSPREE 759
Db 407 VSSSLVDHLRTHHTGYK-----PYKNACEKAYRSCVLT-----QH 442
QY 760 LSEEEEEEDEEEDVDTEDSLARGSESGGKKAISVRGDSSEASGAEEYGVAAATA 819
Db 443 LKTHAAEK-----TSCNACGNSFRN----- 463
QY 820 GKENDSNEKTKQOSSLPPPPPPSLDQDPQMEQSSGVLGKKEGCKPKSPSAPALTP 879
Db 464 --SMCFHRLKTKTKI-----KPYCKDKCKKFTC 491
QY 880 EGEATSVTLVELSLQEARKEPGESSSRKACEVCGQAFPSOALFEHOKTHPEKPLFT 939
Db 492 HSDLTN-----HVRHTGKPKYK--CKEKGKAPRTSSGRIOHLRTHMGEKP-FE 537
QY 940 CVPCRQGF---LERATLKKH 956
Db 538 CDQCGKAFASFSARIHLKTH 558

RESULT 23
US-09-816-669A-14
/ Sequence 14, Application US/09816669A
/ Patent No. US20020137019A1
/ GENERAL INFORMATION:
/ APPLICANT: GARABEDIAN, Michael
/ APPLICANT: TAMEJA, Samir
/ APPLICANT: HITTELMAN, Adam
/ APPLICANT: MARKUS, Steven
/ TITLE OF INVENTION: METHOD FOR SCREENING TRANSCRIPTIONAL COREGULATORY PROTEINS OF
/ TITLE OF INVENTION: TRANSCRIPTION FACTORS, AND ANDROGEN RECEPTOR TRANSCRIPTIONAL
/ FILE REFERENCE: GARABEDIAN-1.1A
/ CURRENT APPLICATION NUMBER: US/09/816,669A
/ CURRENT FILING DATE: 2001-03-26
/ PRIOR APPLICATION NUMBER: 60/225,618
/ PRIOR FILING DATE: 2000-08-15
/ PRIOR APPLICATION NUMBER: 60/191,768
/ PRIOR FILING DATE: 2000-03-24
/ NUMBER OF SEQ ID NOS: 20
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 14
/ LENGTH: 2783
/ TYPE: PRT
/ ORGANISM: Human
US-09-816-669A-14

Query Match 4.3%; Score 229.5; DB 10; Length 2783;
Best Local Similarity 20.2%; Pred. No. 0.00067;
Matches 229; Conservative 117; Mismatches 397; Indels 391; Gaps 50;
```

```

QY 235 PLLPLESPKIPVQISKTLLASSSSSSSSGAETPKO-----AFF 273
      |::|::|
Db 1587 PSOLSHLPLKLPHTS-----TPPOOLANLPOLIPYOCDCKLAFFSFE 1629
QY 274 H-----LNHPLOSQRPFSAAGVGGRSKHPTPARPSRALPGSDIOLABPHILAFSTTGTLAA 328
      |::|::|
Db 1630 HMEOHQDLHFUSAQNOF-----IHPOFLDRSLDMFMLEFDPSNPLLAS 1672
QY 329 QCLGMA-----RGLE--ATASPGLLKPKKGS------353
      |::|::|
Db 1673 QLLSGAIQIPIRASSATSPTSTMTNLKRKLEEKASASBG-----ENDSGTGGEEPRDK 1728
QY 354 -----ELSYGEVM--GPLEK-----366
      |::|::|
Db 1729 RLRTITPEOELELYOKYLULDSNPTRKMHLIDHAHEVGLKKRVYQVMPONT RAREBKQGR 1788
QY 367 ---PGRH-KCFEFCAKVPFSDBALQHLS--HTBERP-YKCNV-----CGRRFTTRG 412
      |::|::|
Db 1789 AVGPQAHRRCFCFRALFEAKTALENHRSIMHNHEAKRACYNUTLSMLDDCGLOMKG 1848
QY 413 NLKVHFHHRREKYPHQMNPBRVEHLDYVTSSGLPYGVSMVPERKAEEEAATPGGVER 472
      |::|::|
Db 1849 DI-----PDGISFS-HLEPPSSD-----GGCVPL 1871
QY 473 KPLVASTALSATESLTLLSTA-----GTATAFGILAEN-KFVLMA---VEPRKNA 521
      |::|::|
Db 1872 SP-VKTMELSPR---TLTSPSIKEVEGIEDFPSPMSVNLNPDQTKRLNDDCSSVNTA 1927
QY 522 DENTPRGSGSASIGVAESSTATMLOLSKMLTSLPSALLTNHFKTSGSPPLPCARLG 581
      |::|::|
Db 1928 ITDTTTGGEGN-----ADDSASATGITETKSSAPBEMGLKRAMMASSEYEDRLSSGLVS 1982
QY 582 ASPSETSKIQOLEVEKIDROGAVALTSAASCAPTTSAPAPSSASSGN-----629
      |::|::|
Db 1983 PAPRYYSK-----EYPNEGTYVDSETSLADPCS-PSPAASGAGKSGBSDGRPOCKRF 2035
QY 630 -----OCVICLRVLSQPR-----ALR 645
      |::|::|
Db 2036 RTOMTNLOLKLKSCFNDRYRTPMLECEVLGNDIGLRKVYQVWFQNAKAEKKSLSMA 2095
QY 646 LHYG-QHGGERP-FKCAKCSGARFSTRGNIRAH-FVGHKASKPAARAONSCPTCOKKFTNA 701
      |::|::|
Db 2096 KHFGINQTSYEERPECTECTLCGIKYSARLSVRDIHFQQHISKVKDTIGSQLDKKEYFDP 2155
QY 702 VTHOOHVNRHNGGOJINGCALPREGGAADAENSEOSTVSGA----GSFPQQSQOPRE 757
      |::|::|
Db 2156 ATVRQLMAQEELRIKANEVL---GLAAOQGFEHDTPRLQALMLPRATVPLQGIPTLL 2212
QY 758 EELSEEEEEDEEDEEDVTD-EDSLAGRSGSESGEKAISYKGSESEASGAEEEVGTAAA 816
      |::|::|
Db 2213 POLNSPSLPGFTPSNALTSPRKNLMGLRPSTYVPSPELLPSGLPNKRSSNASLSPTPQA 2274
QY 817 ATA-GKEMSNEKTTQSSLPRPPPDSDLQOPMED-----852
      |::|::|
Db 2273 TMAMGQQQPPOOQOQOQOPOVOQPPPPRAAOPTPTPLPLQOQOQORXDKSEKYEKEKA 2323
QY 853 ----GSSGVLAGGKEEGGKPERSSPASALPREGATSVTLVEELSLOAMKEP 902
      |::|::|
Db 2333 HKGKEPRLPVPKKEGBAPTATATATISAPRLPYME-YAVDPAQLOALOALATLSDP 2385

RESULT 24
US-09-764-864-954
; Sequence 954, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764, 864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PAIN or file wrapper

```

```

; NUMBER OF SEQ ID NOS: 1792
;
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 954
;
; LENGTH: 708
;
; TYPE: prt
;
; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: SITE
;
; LOCATION: (328)
;
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-954

```

Query Match	4.3%	Score 228	DB 10	Length 708
Best Local Similarity	20.0%	Pred. No. 0.00019		
Matches 183	Conservative 86	Mismatches 326	Indels 320	Gaps 40
QY 108 PTMGPERGSESSGHFLVAATGAAGGGGGLILASPLTGAATPLPEESPAPPPPPPP	167			
DB 13 PLMHPRGRTSQSAAMNASEESFPAGS---	55			
QY 168 PGVSGHILNPLLEELRYLOQROIHOMOTEOICRQVLLGSLQATGAPASPSELPGT	227			
DB 56 -----DHIIC-----GICKQ--QENNIDAFVANHKGSCGLTGT	86			
QY 228 GTASSTPRLPLPSPIKPYQTSKTLASSSSSSSSSSSGAETPRKQAFH--LYHPLGSHQ	284			
DB 87 SAAPTSTQVFESEETVPATQTTQTTTITSETQIT--VSAPEFVEHGYOTPLTPESNEN	145			
QY 285 FSAGGV-----GRSHKPRPARSP-----ALPGSHDQLASPHLAFSTTGLLAOCIGAA	334			
DB 146 QTATVATSLPAKSRFKKPTTPPAOKRYNCCYR-----COFKTIYGM-----	186			
QY 335 RGLSATASPGILKPKNGSGELSYGEVWGPRLKPGRHKRCRCAYFGSDSALOIHLSRHT	394			
DB 187 KDMGRH-----LKITGT-----DKP--HKCEYCGKCFSRKDKTKTHMRCHT	225			
QY 395 GERPYKCNVCGNFTTTCGNLKVHFHRRREKYR--HYQMPHPRPHELDYVITSSGLEPGMS	453			
DB 226 GVKPYKCKTCQDYAAADSSSLNKLRLTHSDRPFQICQIPY-----	265			
QY 454 VPRPEAEETAPPGGVGEKKPLVASTTALSTATELT--LTSAGATAPAGLPAFNKFLM	512			
DB 266 -----ASNNSQLTVHLNHSHTGAPPOCMCSAKFKIS	298			
QY 513 KAVEPKNADEWTPPGSESGAISGVASESTATLQMLKMTSLPWWALTNNHK--STG	569			
DB 299 SDLRHHMVHSEKRFK-----CEFCNVRCITMK--GXHKSHIRIKHSGN	340			
QY 570 SFPLPLCARALGASPESTSKLOQLEKIDROGAVAVTSAASGAPTTSAAPSSASSGPN	629			
DB 341 NFKCPHC--DFLGDASKATLRKHSRVHQ-----SEHPE	370			
QY 630 QCVCICLRVLSRCRALRLHNGHGERPFKCVCGRAPSTR--GNLRANP-----	676			
DB 371 KCSECSYSCSSKAALRIHERIHTDPRKYNIC--SEPTKPSNLSKHKKTFHGDWVKT	428			
QY 677 -----VGHKASPA-----ARAONSCPICOKKFTNAVYLOOHVRMH-----L	712			
DB 429 ALERKDTGRQSSROYAKLADAKKSFHCDICDASFHREDLSRBHKRHQSEYSEKNSVTVL	488			
QY 713 GGGI--PNGGTALPEGGAQAQENGSEOSTVSGAGSFPOOOSQOPSPBEELISEEBEEDDEE	771			
DB 489 QFQIDPSKOPATPLTVGHLO-----VPLDPSQVP--QFSE-----	521			
QY 772 EEDYTDSESLAGRGSSGGEKAISVRKDSSEASGAEEVGVYAAATAGKEMDSNEK--T	829			
DB 522 -----GRVKTIYHQ--VPQANTIVQAAAANAIVIPALVAQNDEELPGNSRLQI	569			
QY 830 TQSSSLPPPPPP-----SLDQOPM-----EOGSSGVL-----GKGECKGRPER	869			
DB 570 LROYSLIARPGQSRCPSEKAGAMOTOPAVLLTHEDTGDATLTIQTLPLTAGGQPEGSGNOT	629			
QY 870 SSSPA-----SALTPEGEATSVLVEE--LSIQEAMNKEGESSSRACEVCGQA	917			

```
DB 630 FITSSGTCITDCEGLNALIQGCTA-EVTIVSDGGQNIIVATTAAPVSSSSSQ-----QE 682
QY 918 FPGQALIEHQKTHP 932
DB 683 LPKQTVSIIQGAHP 697

RESULT 25
US-09-764-864-947
; Sequence 947, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR APPLICATION DATA: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 947
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-947

Query Match
Best Local Similarity 24.4%; Score 227.5; DB 10; Length 323;
Matches 90; Conservative 45; Mismatches 133; Indels 101; Gaps 16;

QY 365 EKPGRHRCRCFCAYVFGSDSALQIHLRSHTGERPYKCNVCGNRFRTTSGNLKVFHRRREK 424
DB 11 EKP---HRCILCPASAYERHLEAHMRSHTGKPYKCELCSPFCSDRSNLS---HHRRRK 64
QY 425 YPHVQNM-----RPHVPEHLDYVIT-SSGLP 449
DB 65 HKMVPKIGTRSSLSKKMWGLQKKTSLNIGYSRRALINLSPPMVYOKPPLYLNDFTHEIP 124
QY 450 YGMSVPEKAESEA-ATPGGVERKPLVASTALSTESITLSTSGTATAPGLPAFNK 508
DB 125 ---NIQTDSTESMAKTTPTGGLPRD-----QELMVDNPLNQLSTLAGQLSS--LPP--- 171
QY 509 FVLMKAVEPKNKADENTPPGSEGSALISGVAESSTATLMQLSKLTSLPFWALLTNHFKST 568
DB 172 -----ENQNPASPDVYPCDEKPFM-IQPSQAV--VSAVASASIPQ-----SSS 213
QY 569 GSFPPLPCARLALGASPSSETSKLOQLVEKIDRGAVAVTSAASGAPTTSAAPSSSSASGCP 628
DB 214 PTPSPER-----RPSHSORNYSPV-----AGPSEPSAHTSTPISGNSQSTPAPALP 260
QY 629 NQ-----CVICLRVLSCPRALRLHYGONGGERPFKCKVCGRAFSRGNLRAHFVGHKA 681
DB 261 VQDPQLLHHCQHCMDYFADNLIYTIHMGCHGYENPQCNICGCKKKNKYFACHF----- 315
QY 682 SPARAONS 690
DB 316 ---ARGQHN 321

RESULT 26
US-09-764-864-1095
; Sequence 1095, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR APPLICATION DATA: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 1095
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-1095

Query Match
Best Local Similarity 24.4%; Score 227.5; DB 10; Length 323;
Matches 90; Conservative 45; Mismatches 133; Indels 101; Gaps 16;

QY 365 EKPGRHRCRCFCAYVFGSDSALQIHLRSHTGERPYKCNVCGNRFRTTSGNLKVFHRRREK 424
DB 11 EKP---HRCILCPASAYERHLEAHMRSHTGKPYKCELCSPFCSDRSNLS---HHRRRK 64
QY 425 YPHVQNM-----RPHVPEHLDYVIT-SSGLP 449
DB 65 HKMVPKIGTRSSLSKKMWGLQKKTSLNIGYSRRALINLSPPMVYOKPPLYLNDFTHEIP 124
QY 450 YGMSVPEKAESEA-ATPGGVERKPLVASTALSTESITLSTSGTATAPGLPAFNK 508
DB 125 ---NIQTDSTESMAKTTPTGGLPRD-----QELMVDNPLNQLSTLAGQLSS--LPP--- 171
QY 509 FVLMKAVEPKNKADENTPPGSEGSALISGVAESSTATLMQLSKLTSLPFWALLTNHFKST 568
DB 172 -----ENQNPASPDVYPCDEKPFM-IQPSQAV--VSAVASASIPQ-----SSS 213
QY 569 GSFPPLPCARLALGASPSSETSKLOQLVEKIDRGAVAVTSAASGAPTTSAAPSSSSASGCP 628
DB 214 PTPSPER-----RPSHSORNYSPV-----AGPSEPSAHTSTPISGNSQSTPAPALP 260
QY 629 NQ-----CVICLRVLSCPRALRLHYGONGGERPFKCKVCGRAFSRGNLRAHFVGHKA 681
DB 261 VQDPQLLHHCQHCMDYFADNLIYTIHMGCHGYENPQCNICGCKKKNKYFACHF----- 315
QY 682 SPARAONS 690
DB 316 ---ARGQHN 321

RESULT 27
US-09-764-864-893
; Sequence 893, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR APPLICATION DATA: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 893
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-893

Query Match
Best Local Similarity 17.9%; Score 226.5; DB 10; Length 361;
Matches 108; Conservative 43; Mismatches 161; Indels 291; Gaps 15;

QY 365 EKPGRHRCRCFCAYVFGSDSALQIHLRSHTGERPYKCNVCGNRFRTTSGNLKVFHRRREK 424
DB 9 ERP---PECEBEGKAFREKSTVIYIHRTHTGKPYECNCGKAFYOKSNLIYHQKTH--- 62
QY 425 YPHVQNMHPVPEHLDYVITSSGLPYGMSVPEKAESEAATPGGVERKPLVASTALSA 484
DB 63 ----- 62
QY 485 TESTLTLSTAGTATAPGLPAFNKFVLMKAVEPKNKADENTPPGSEGSALISGVAESSTAT 544
DB 63 TGKTYECTKCG-----ESFI-----OKLD----- 82
```



```

722 ALPEG 726
Db 450 SPROG 454

RESULT 31
US-09-864-761-36148
; Sequence 36148, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36148
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC003682.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.6

```

```

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALU0 1.00e-104
; OTHER INFORMATION: SWISSPROT HIT: P52740, EVALU0 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALU0 1.00e-104
US-09-864-761-36148

```

Query Match	224;	Score	224;	DB	10;	Length	536;
Best Local Similarity	23.6%;	Pred. No.	0.00022;				
Matches	115;	Conservative	42;	Mismatches	180;	Indels	150;
						Gaps	21;

OY	273	FHLVPLDGGHFFSAGGVGSHKPTAPVSPALGSDULLASP-----NLFFSFUGLLA	326
Dd	107	FHV-----SGKFPTFEVGDFGSAT---SLILOHVPTTERPHSRILNHLVPT-----	152
OY	328	AOCCLGARBLELATASPOLLKPKN-----GSGE-----LSYGEVMGPL--EKPG	368
Dd	153	----GRKLAKTESKRSSFREKSVFLOHORADSGERPKYCECGSKFSQSQGFLLHRRAH	207
OY	369	GR--HKCRFCAKAVFGSDSAIQLHLRSHTGERPKCVNCVGNRFTTGNLKVHHRRREKY	425
Dd	208	GTRTHHECSECGKSFBRKHTLHQVRVHTGERPYDCSECGKSFROVSYLIOQRVHTGER	267
OY	426	PVVOVNPHPVPEHLDYITTSGLPYGMSVPERKEDEELATGGVEREPVASTALSAT	485
Dd	268	P-----YESCGKSFHSHTMLYR-HRSMT-----STRPYECSECGKFS	307
OY	486	ESLTTLSTSGATPAQLPANKFVLMAKAVEPRKKADENTPPGSEGSAISGAESSSTATL	545
Dd	308	HSTNL-----FRHHRVHTGVARPCSE-----CGKAFCNTLYL	340
OY	546	MOLSKIMT-----SLDSWALLTNHF-K-STGSFLPLCKRALGASSTFSSKLOO	592
Dd	341	IHHORFHTGERPVYCECGKSFQOKSVLIOQRVHTGERPYE-C-----SECCK---	388
OY	593	LVEKIDROGAVATSAAHG-----APTASAPASSASSSPNOVCICULRYLSCERALL	646
Dd	389	-----VFSQSSGLEFRHRRAT-----KTMRPYESECKSPCKDILLR	426
OY	647	HYGOGHGERPFCKVCVGRASFSTGNLRHAHYGHKASPARAONSPICOKRTTNVATLOO	706
Dd	427	HOVHTVHGERPYECGSKFSFR-----KTHLIRHQTVHTNERPYECDECGKSYSQSALLQ	482
OY	707	HYVMMHG 713	
Dd	483	HRRVHTG 489	

```

RESULT 32
US-10-037-667-1
: Sequence 1, Application US/10037667
: Patent No. US20020177145A1
: GENERAL INFORMATION:
: APPLICANT: Morgan, Bruce A
: TITLE OF INVENTION: REGULATION OF NEURAL DEVELOPMENT BY
: FILE OF INVENTION: DADA10S
: FILE REFERENCE: 10287-044001
: CURRENT APPLICATION NUMBER: US/10/037,667
: CURRENT FILING DATE: 2002-07-23
: PRIOR APPLICATION NUMBER: 60/243,110
: PRIOR FILING DATE: 2000-10-25
: NUMBER OF SEQ ID NOS: 13
: SOFTWARE: FaastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 537
: TYPE: PRT
: ORGANISM: Mus musculus
US-10-037-667-1

```

Query Match	4.2%	Score 224	DB 9	Length 537
Best Local Similarity	23.1%	Pred. NO. 0.00023		
Matches 137	Conservative 45	Mismatches 222	Indels 188	Gaps 25

QY 227 TGTASTKPLPLPSPKP-VQTSKTLASSSSSSSSSGAETPKQAFHLHPGSHPF 285

QY	216	GAPASPEL-----	PGSTASSTRPLRLPRSP--	IKPVOT	248
Db	213	GPKLPKPSASABGNISVVOQLGTLPHNALQLSACPAPVASAGSGSTLPAFEGPOHLSOASG			272
QY	249	SKTLASSSSSSSSSGAETPKQAFPHLYPLGSOHP-----	FSAGGVORSHKP-----	TP	298
Db	273	TSTCCSTISAAPRDSGA-----	HPACSTGPBARCAVAASSTVGNAAVQPOANSP		321
QY	299	-----	APBALPGSTDOLIASPHLAPRPTTGLLAOCLGARGLEATASFG		344
Db	322	PALGPRLLSSASNLPNLLPOT-----	SSSVIPEPNPLVSTIA-----	TANALDPLISA--	370
QY	345	LLKPRNGS-GELSTYEWGPRLEKPEGGRKRCRCACAVFGSDSALOHLRSHNTGPRPKCV			403
Db	371	LKMHKGRPNVSYVEPRASADDPFFKHKRCRCACAVFGSDSALOHLRSHNTGPRPKCV			430
QY	404	CGRNRTYGNLKVHNRHREKYPHYOMNPHVRENHLYITSSGLPYGASVPEK-----			458
Db	431	CGRNSTYGNLKVHNRHREKYPHYOMNPHVRENHLYITSSGLPYGASVPEK			490
QY	459	AEEBATP---GGGVERKPLVASTTALSTATESLTLSS-----	TSAGTAVAGL--		503
Db	491	LDSKVLPTVPTSVGLQLPRTVPGTHNYTDSITFVSSRQPRSPASSRCTSLSPGLNN			550
QY	504	-----	PAPKTVLKAYPERKKADENTPRGSE--	GSALISGVAESTATIMQ	547
Db	551	TESGITVRESPOPLRLLGSRPLTKA--EPVSLPCTSTTGARPVUGGVGSGLPJTAATAVTD			609
QY	548	LSKIMTSLPMSALLTNHFKSTGSPFLPLCARALGASPESTSKLOOLVEKIDROGAVAVTS			607
Db	610	SACTSLGSBGLPRAVDQFKA--QFPRGGLLDSH--QTSESKILOOLENIDK			658
QY	608	AASGAPTTAPARSSASSASSGPNOCVITCLVYLSCPRALRLHYGONGSERPREKCVCGRAFS			667
Db	659	-----	MTDPMQCVICHVULSCQSALIKMHYPTNHTGPRPKKICIGRAFT		701
QY	668	TGRLNLRANHVGHKASPAALAAQNSCPLCKKFTPAVYLOOHVKNHNLGQITNGGTALPRGG			727
Db	702	TGKNLKTHGVHGRKPLRVHOHSCPLCKKFTPAVYLOOHVKNHNLGQITNGGTALPRGG			759
QY	728	GAA-----	QENSSEOST-----	VSQAGSFPQOQSO	752
Db	760	QEMAMADLPFDEKNAETLSSFPDDIDENSMEDBSLKDTPASDSKPLLSVSGSCP-----			814
QY	753	QSPPE-----	EELSEEEEBEDEEBEDVYDEDSLACRG--	SE	787
Db	815	PSPPEVSIISIALENOQMKMIDSVNMQOOLANLKSVENGSGESDRLSNDSSSAVGDELSR			873
QY	788	SGGEKAIVSRGDEEASGABEEVGYVAAAATA--GKENDSNEKTTQOSSLRPPRPDSDLO			846
Db	874	SAGSPRLSSSSSSQALSPHNSNGESRFSRSPGLGHOEDPDELKTERLDSPP-----			926
QY	847	POPMEGSGSGVLGKREBGGKP--BRSSSPASALPRREGATSVTLVEELSLQEDAMKEGES			905
Db	927	PGRGNGALDITAG--HGRPLIKEARPSLL-----	FLSRENGKC		965
QY	906	SSRKACEVCGOAFPSQALAEHOKTHPKREGLTTCVFCROGFERATLTKNMILANHOVQ			965
Db	966	AS--TVGCVGCGKPRACKSALEIHYRSHTKE--RRFVCTVCGRGCSITMGMLKQILTL--HKLK			1021
QY	966	PFAHPHQPONIALSLVPGCSPTITSTGSLSPF			997
Db	1022	EL-----PSOVFDENFTLGPBHSHTPSLASSPAP			1049

RESULT 3
JC7116
Xsal-3 protein - African clawed frog
C.Species: *Xenopus laevis* (African clawed frog)
C.Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-May-2000
C.Accession: JC7116
R.Onuma, Y.; Nishinakamura, R.; Takahashi, S.; Yokota, T.; Asashima, M.

Biochem. Biophys. Res. Commun. 264, 151-156, 1999
A:Title: Molecular cloning of a novel Xeropus spalt gene (xsa1-3).
A:Reference number: Jc7116; MGID:99458636; PMID:10527856
A:Accession: jc7116
A:Molecule type: DNA
A:Residues: 1-1061 <NU>
A:Cross-references: DDBJ:AB030827
A:Experimental source: egg, neural tube
C:Genetics:
A:Gene: xsa1-3
C:Keywords: egg; zinc finger

Query Match	21.7%	Score 1145;	DB 2;	Length 1061;
Best Local Similarity	31.5%	Pred. NO. 1.2e-47;		
Match 324: Conservative	140;	Mismatches 347;	Indels 218;	Gaps 36;

[illegible]

A:Gene: FlyBase:DvIr/seIm
A:Cross-references: FlyBase:FBgn0013337
C:Keywords: DNA binding; Transcription regulation

Query Match 14.1%; Score 745.5; DB 2; Length 1402;
Best Local Similarity 22.1%; Pred. No. 2.3e-28;
Matches 322; Conservative 134; Mismatches 383; Indels 619; Gaps 52;

```
OY 2 AHESESSRLGAVGAEELGADSEEDH-----POYCAKCCAOPTPTPEFLAHONAC 54
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 57 ANASAGNSPCPSRSRPOQSEDEREPOVSEQLVPEVSAOSESEIGEIE---NNAD 112
OY 55 SDPPRVAVIIGGQENPNSSASSEPREGHNNPOVMDTEHSNP-----DSGSSVP 105
D 113 ETN-----ADHNHNN-----NNKLYM-----TPRVEHEVQNNANLNASMP 150
OY 106 TDPYWPERRGESESGHFLVAATGTAAGGGGLLAPKLGATPLRPPESTAPRPPPPP 165
D 151 NSTY-----PRATNAVIGA-----RAQGFATPTVLEAIQNMOMALAQF 190
OY 166 PRPGVSG-----HLNPLLEEELRVLAQROIHOMQMTQICROVLLG----- 209
D 191 AAKTIANGASGTDEAAMKQALFLQOTLFNLQOQDLFQQLLQQLSOLALNOVKQNDDE 250
OY 210 -----SLGOTVGAAPSPSE-----LPGTGASSTKPLP 238
D 251 ADELEPEREEDGETDYEEERLADMELROKAEAMASAKOHILNAGVRYANAP--DP 309
OY 239 LESP-----IKPVOTSKTLASSSSSSSSSG--AETPKQAEFHLVHPLGSOHPFA-- 287
D 310 SHQPRHRCRLRLKRRKEEDASAKSSGASAKTGEQESSQDALNKLKEMENMPPLFGADL 369
OY 288 -GGVGSNKHKTTPAS--PALGSGTDQLASPLHAPSTGTGLAAQCCGARGLEATSPG 344
D 370 SSSIITNHDLLPEPNSLDLQKRTQVELDS-----ASQG 403
OY 345 LKPKNGSGELSYGEVMGP---LEKPGGRHKCFCAKVFSGSDALQTHRSHTGERPKC 401
D 404 IL-ANNMADDEFAPDSSDGKGRNRPFRKHRCYCGKVFSGSDALQTHRSHTGERPKC 462
OY 402 NVCGNRTTGNLKVHNRHREKRYHVOMNPRVREHLDV-----I 443
D 463 NVCGSRRTTGNLKVHNFQRHAKFRPHVPMNAPIREHMDFNRPRLDOMSPSSPTQSPA 522
OY 444 TSSGLPRYGMSPREKA-----EEEAATPG--GGVARKPLVASTTALSAETSLTL 491
D 523 PATGLP-----PPESTLTLTQMPSSSFASSPAFPGLRGTIRPPY-----ELKSL 567
OY 492 STSAGTATA-----GGLPAPFKFV-----LAKKAVEPKKKADE-----NTP 526
D 568 GATAGTAGLRLPRFFRPMRGIGALAKHTHDQSMDPTDLKSSGSPSRPHEEDNIAARLP 627
OY 527 PGSE-----GSAISGVASSTATL-----MO-----LSKLM 552
D 628 VKSELMEEKTEHTMEATRESAEEMERPLLEVRIKEREIDEDQMHLQEGMQKREPLTAY 687
OY 553 TSLPSMALL--TNH--FKSTGSPRLPLCAR--ALGASPSETSK-----LQQL--- 593
D 688 TPHPQOCLITPTTHAAKSPRSLPLQOCHARLSLWCSHPRTSNHACAVLTGSOITHLQDLTP 747
OY 594 -----VEKIDROGAVAVTSAASGA-----PTTSA 617
D 748 DNVPTPMQOREDFEAFERPLNTSKTDHSPISRPAGNANAHIRSPREFNFIKHEMAFV 807
OY 618 PAPSSAS-----SGPNOCYTICRLVLSCPALRYLHGOH 651
D 808 PRPNSNDMSWENFIEVSNSETMKLKLKLNKKKISDPNOCVCDRLVLSCKSALOMNHYRTH 867
OY 652 GGEPRFKCYGGRASTREGLRAHFEVGHKASPARAARONSCPIQOKKFPNAVATLOOHVNRH 711
D 868 TGEPRFKCRICGRAFTTKGNLKTTHAAVNHKIRPRMRNHOCPVCHKKYSNALVLOOHIRH 927
OY 712 LG-----GQIPNGGTALPEG-----GG 728
```

```
D 928 TGEPTDLPQIQAAETIRDPSPMPCGHFMNPFMAAAPHGAMPGGAGGPRPATGMGG 987
OY 729 ----- 728
D 988 PHNGTIGSESSQGLDDNMDCGDDPDFDISSEHLNSNDPAATSDRRSSDDFKSLLEQ 1047
OY 729 -----AAONGSEQSTVGAGSFPOQSOQP-----SPEE 758
D 1048 KLRIDPTGVYVNIHQHRSAAASNNSIGSASASBPAPSPSQPKRPSQVRSQSPYR 1107
OY 759 ELSEEEDEEEDYDTE---DSLGRGSESGEKALISVG----- 798
D 1108 SVS-----ETSGALDLTPRALPRPLASSSSRSRPYQLLSVRRRLARSYSHRCVPMV 1162
OY 799 ----- 798
D 1163 RALLSQLPSPVGIDCLPRGLQNLQOQHNLMOQOXAVAAALAAQNNHNMOMQAAALH 1222
OY 799 -----DSEASGAEEVGTVAATAATAGKEMPSNEKTTOQSSLRPPRPPLSD 845
D 1223 QHOEHLREAOEVQKAAQEVQKAAAAAARQES-----PQRPFRS-- 1269
OY 846 QPQMEQSSGYLGEKEGKPERSSPASALPEG-----EATSVTLVEELS--L 894
D 1270 -----GESSVGPAPQPNPLISARPPFGMFPMPLPIFRPATTOCMNAMNOIA 1315
OY 895 QEAMRKEP-----GESSKRACEYCGAFPSQALAEHQNHKPKRPLTCVFCRCGF 947
D 1316 QSVNPAAPFNPLAISGVNGS--TTGCTCYKTPCHSALEIHYRSHTKERP--FKNICDNGF 1373
OY 948 LERATLKKHMLAHNOVQ 965
D 1374 TTKGNLKKQHMILT--HKIR 1389
```

RESULT 6
T13805
spalt-related protein - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13805
R:Barrio, R.; Shea, M.J.; Carulli, J.; Lipkow, K.; Gaul, U.; Frommer, G.; Schuh, R.;
submitted to the EMBL Data Library, August 1996

A:Description: The spalt-related gene of *Drosophila melanogaster* is a member of an
A:Reference number: Z1768
A:Accession: T13805
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1263 <BAR>
A:Cross-references: EMBL.Y07653; NID:e1004366; PID:e267584; PIDN:CAA68937.1
C:Genetics:

A:Cross-references: FlyBase:FBgn0000287
A:Introns: 862/1; 915/3; 1003/1; 1080/3; 1146/3
Query Match 13.8%; Score 727; DB 2; Length 1263;
Best Local Similarity 23.5%; Pred. No. 1.6e-27;
Matches 275; Conservative 144; Mismatches 397; Indels 356; Gaps 39;

```
OY 67 QENPNSSASSEPREGHNNPOVMDTEHSNPPDSGSSVPTPTMGPERGESSGHF--- 123
D 38 REIRDQTLAKELDADSNNGTEPOMEAEAVPESDTERETAEREGEREPENSNEALDIS 97
OY 124 LVATGTAAGGGGGLT-----LASPKLGATPLRPPESTAPRPPPPRPGVSGH 174
D 98 LISSGRESLPSSGHSVLEALQHTKVAVAPATAMAGNHQSA----- 140
OY 175 LNIPLLEEELRVLAQROIHOMQMTQICROY---LLG---SLGOTVGAAPSPSELPGT 227
D 141 -DLAVQSTIFNVQRQHNMQDLQIQLHOSQKRAEAAALGHSNHSDEDEEEREPEPEPK 199
OY 228 GTASSTKPLPLFSPIKPVOTSKTLASSSSSSSSSGAETPKQAFHNLVHPLGSOHPFS- 286
```

Db 200 QPTNGLEKELEEQPESEODHES- RREENSKTDKRGTEDRKA----- EREGQSMCD 252
QY 287 -----AGVGRSHKPTAPR----- PALPGSTDQLASPHLAPFTTGLLAQCGLGARG 336
Db 253 ISSSLASSITTNDDPPAPNEPCLEMLQRTREYLD----- ASOSLHAQO- 299
QY 337 LEATASPGLLKPKNGSELSTGEVWGPLEKPGGRHKRCRCAKYFGSDSALOIH-RSTTGE 396
Db 300 MGEVSEYASKESKQSRREIF----- KIRCKYCGKIFGYSALQIHLRSHTGE 346
QY 397 RRYKVCVGNRRFTTRGNLKVHFHREKYPHYQMNPHVPEHLDVYTSSG----- 447
Db 347 RPFVNCVCSKFTTKGNLKVYGRHTQTFPPMLLPBGVAPN----- YGSHSGQGVQGEQY 401
QY 448 ---LRYGMSVPRPKAE-----EEAATPGGVKRPVLVASTTLSTESTLSTSA 495
Db 402 PIRLPAPRYAPVGOEHQNOVEEPEELPVOAEDLSKPVKEKE--KSHSPVECVKTPK 459
QY 496 GTATAPGLPAFNKFLMKAVEPKNKADENTPPGS-----EGSAISGVASSTATL- 545
Db 460 EVKTDASLPSSSEK-----PEKEISKPVYTSRRNGSVKRRQTSVAPRQEDRENDIV 511
QY 546 --MOLSKMTSLPSMALLTNHFKSTGSPFLPLCARALGAS-----PSETSKIQOLV 594
Db 512 EHMNTAKLVRSSA-----SRESOPAEYSIAQMERIDKSWEDLIEIDKTSSETSKLOOLV 566
QY 595 EKIDROGAAVAVTSAASGAPTTSAPAPSSASSGPNQCVICLVLCSPALRLHYHGOHGE 654
Db 567 DNIENK-----LTDPRQCIFCQKVMSCRSSLOMIRHTTGE 602
QY 655 RPFCKVCGRASTRGNLRAHFVGHKASPAARAQNSCPICOKFTNAVTLQOHVHML- 712
Db 603 RPFRCIKCRAFAETKGNLKAHMSIHKIKPQRSQFCPCVHQKFSNGIILQOIRIHTMD 662
QY 713 ---GGO-IPNGGTALPEGGAQENG-----EGSYSGASSFPQ 748
Db 663 DGSQGGVVAANPGEABRUGIEDONSNKSLGTSPTLDFSTTISDHSGQSESSQGDGDFDE 722
QY 749 QOSQO-----PSPPEELSEEEDEDEEEDVTDSDS-----LAGRSES 788
Db 723 FMTWSTDSDRONSAAATATPHPLERERDRKERRIRIPNOCSDERSHNPDLTGRRSES 782
QY 789 GGEKASIVAGDSEA-----SGAEEVGT 812
Db 783 GEMPRAMDLSPPSSNSGRIFATGLANGATGGSGNGGLPMLGMPMPNLLILMAAREEMHA 842
QY 813 VAAA-----AT----- 818
Db 843 LGHAHAKFLLPFGPLGFMGLHPPRNVNCNLCFKMLPSLALESILQSEHAKERATGHAOR 902
QY 819 -----AGKEMDSNEKTTQOSSLP-----PPPP--DSLD 845
Db 903 PHCSADGSPYGAKLTLNPLMFAKKPSSSSSSSGEKLPESSNPPEAPMPATPIKEDPD 962
QY 846 QPQPV-EGSGSEVLGKKEGKERSSPASALTPEGEATSVTL-----VELSLQEA- 897
Db 963 QEQLVVERBAS-----AGESGTGATSNTPQEDAGDEQSLMKQHLAHNRPASPLDFQDAL 1018
QY 898 MKREGEES-----SRKACEVCGOAFPPSOALEEHQKTHPKEGPLFTCVFROGLEEA 951
Db 1019 MSAGPPTSLDPPVNNKHCHVCGRNRNFSSSSALOIHMRTHIGDKP-FOCNVCOKAFPTTNG 1077
QY 952 TLKKHMLLAHHOVOPAPRPGPONTIAALSLVPG 983
Db 1078 NLKVHM-GTHMNTNPTSRGRRMSTLELPMRPG 1108

RESULT 7
20969
Hypothetical protein F15C11.1 - Caenorhabditis elegans
Species: Caenorhabditis elegans
Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
Accession: T20969

R.Wilkinson, J.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19352
A:Accession: T20969
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-744 <NT>
A:Cross-references: EMBL:Z19352; PDB:1CA95798.1; GSPDB:GN00019; CESP:F15C11.1
A:Experimental source: clone F15C11
C:Genetics:
A:Gene: CESP:F15C11.1
A:Map position: 1
A:Introns: 26/1; 75/3; 108/3; 145/3; 251/1; 453/2; 500/2; 613/1; 649/1
Query Match 7.48; Score 391.5; DB 2; Length 744;
Best Local Similarity 21.88; Pred. No. 1.1e-11;
Matches 194; Conservative 89; Mismatches 278; Indels 327; Gaps 32;
QY 219 ASPSELPGTGTASTKPLPLPSPIKPVQTSKTLASSSSSSSSSGAETPKQAFPHLYHP 278
Db 18 SKRRMSGEEDA-----MMSPL-----DLSTKSFDENCKE----- 48
QY 279 LGSQHPSAGGVGRSHKPTAPSPALPGSTDQLASPHLAPFTT-----GLTAAQCLGA 333
Db 49 -----GAGG-----ALP-LEDRSNILLPHFSVPFANPQQLSLCAQLGNSS 87
QY 334 ARGLEATASFGLLKPKNGSELSTGEVWGPLEKPGGRHKRCRCAKYFGSDSALOIH-LRS 392
Db 88 SRNVSTASTTSSCP-----IQSCSOSFSSPAAITWHVLA 123
QY 393 HNGERP-YKNCVGNRPFTTRGNLKVHFHREKYPHYQMNPHVPEHLDVYTSSGLPYG 451
Db 124 HDEQGLFSDVCTTTS-----NGQDIRH----- 149
QY 452 MSVPRPKAEEREAATPGGVKRPVLVASTTLSTESTLSTSAGTATAPGLPAFNKFLV 511
Db 150 -----KCOQTLASRSTSVPTTIPSSVCFLS-----TPTPCL-----QFSL 186
QY 512 MKAVEPKNKADENTPPSEGSALSGVAESSTATIMQSLKMTSLPSMALLTNH--FKS 567
Db 187 NSIGTSETREDEEDMDVEDGEHVANQLFGHLQKSDSKMS--LFNAPRPFPA 243
QY 568 TGSFPLPCAR-----ALG-----ASPSETSKLOOLVEKIDROGAV 603
Db 244 FPNMPPEPLMRQGFDPRAVFAGRHNDNDDEALMELSTSEAKIRLIV-----GDK 297
QY 604 AVTSAASGAPTTSAPAPSSASSGPNQCVICLVYLSCPALRLAHYGQGEKPFCKVCG 663
Db 298 AV-----PTTD-----PNOCTLCRYLSCKSLQOMHYRTHTGERPFCKICQ 339
QY 664 RAFSTRGNTRAHFVGHK-----ALG-----ASPSETSKLOOLVEKIDROGAV 603
Db 340 RAFPTTKGNLKTGHGVRKSKSPFGRLPSLPPQLAAMHONOHQIAPPRHINHPITSAAS 399
QY 663 PAA-----RAQNSCPICQKFTNAVTLQOHVHMLG-----QIPNGSTA-----LP 724
Db 400 AAAAVALQIASOQCPICQGFNLNAGELAVHTEHRNSLTQPRVAPRTPTTTRVQTFEVP 459
QY 725 EGGGAQENGSEST--VSGAGSPFQOQSQSPSPPEELSEEEDEDEEEDVTDSDSLAG 783
Db 460 FTTTPPSLNTADSTQFNLANLISLAQLKNDSSPNTDVSSVEEKTTRD-----DPPKMS 513
QY 784 RGSSEGGKALSYRGDSEASEEVEGTVAATAAGKEMDSNEKTTQOSSLPPPPDS 843
Db 514 LSPNSSDSSSVRODILSESEFEKL-----KLEPEPLLEQVSTTPNP----- 559
QY 844 LDQOPMEQSSGVLGKKEGKERSSPASALTPEGEATSVTLVELSLQEAAMKEGK 903
Db 560 -----KNEPDLAMQKMAATEPPRQMPV----- 585
QY 904 ESSSRKACEVCGOAFPPSOALEEHQKTHPKEGPLFTCVFROGLEKATLKKHMLLAHHQ 963
Db 586 --LSKHQCGVCFKHFSSSSALOIHMRTHIGDKP-FKDCMGRALFTTRGNLKVHMGTHSWQ 642

964 VQPF-----APRGPONIALSLVPGCSF--SITSTGLS 994
 Db 643 QSPSRGRRIQFDVASSVTEKPMLOSPILPTSGAGASPLAMLGPNGLS 690
 RESULT 8
 742717
 DNA-binding protein Rc - mouse
 A:Alternate names: Ig kappa chain gene enhancer Recognition component
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
 C:Accession: T42717
 R:Wu, L.C.; Lin, Y.; Strandmann, J.; Mak, C.H.; Lee, B.; Li, Z.; Yu, C.Y.
 Genomics 35, 415-424, 1996
 A:Title: The mouse DNA binding protein Rc for the kappa B motif of transcription and for
 ew family of large transcriptional proteins.
 A:Reference number: 22238; MUID:97001141; PMID:8812474
 A:Accession: T42717
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2282 <WU>
 A:Cross-references: EMBL:L46815; NID:91377885; PID:91377886; PIDN:AA040884.1
 A:Experimental source: strain BALB/c; clone T1; thymocyte, brain
 C:Genetics:
 A:Gene: Rc
 C:Function:
 A:Description: binds V(D)J recombination signal sequence and kappa B motif
 C:Superfamily: HIV-Ep2 enhancer-binding protein
 C:Keywords: DNA recombination; transcription factor
 Query Match 6.6%; Score 346.5; DB 2; Length 2282;
 Best Local Similarity 23.0%; Pred. No. 5.2e-09;
 Matches 224; Conservative 114; Mismatches 336; Indels 299; Gaps 44;
 203 RQVLIGSGQYTGAPASPELPGTGA-----SSTKPLP---PLFSPKPVOTSKTLASS 255
 18 RRRLTGEMAIQTIVSSSAP--YPSGTTAPESATQELLATQPFSP-----SQ 64
 256 SSSSSSSSGAETPK-QAFPHLYHPLGSOHPVSAGVGNSHKPTPA-----PSPALPGST 308
 65 EKTGOQKPRARRPSIEASVHT-----SOLP-----QHPLTPAFMSPGKPEHLLEGST 111
 309 DQLI-----ASPHIAF-----PSTTGLLAQCL-GAARGLEATPASPGLTKKNG 351
 112 WQLVDPMRPEPSSSEVAPGSHQSQLPSHASILPPEELPGIKVVPVPSQVSLKP--- 168
 352 SGEISGEVWGPLEKPGGRHKRCFAKVFSDSALOIHLSHTGERPYKCNVGNRFETTR 411
 169 -AEAAHKKEKRP-QKP-GKYICQYCSRPCAKPSVLOKHIRSHNGERPYPCGPGSFKTK 225
 412 GNLKVFHHRHREKYPHVQMPDHPVPEHLDYITSSGLPYGMS---VPPEKAE----- 461
 226 SNL-----YKHKSHAHRIKAGLASGSSSEMYPPGLEMERIPGEFEETPEGEST 275
 462 -----EATPPGGVGVERKP-----LVASTALASRESTLLTSAGTA---TPGL 503
 276 DSEETEGAAAGSPSTDVLPKPNHPLSSSLYSIGSSGSSQERCSLSOSTPSPLEDPAPFA 335
 504 PAFNKFEVL-----MKAVEPKNKADENT--PPGSEGAISGVAESSTATLM 546
 336 EASSEHPLSHKPEDTHTIKQALALRSEKKLIEBQTFLSPGKSTESGVFSSESSEAEQ 395
 547 QLSKMTSLSPSMALL-----TNHFKSTGFRP-LPLCAR---ALGASPSSESKL 590
 396 QVSPPTNINASTYAEIIEGCGRIGRTSMALASTSQPLPLPSSEDKPSLVPLSVPRTOVI 455
 591 QQAVEKIDROGAVAVTS-----AASGAPTTAPAPSSS----- 623
 456 EHITKLTITNEAVVDTSIDSVKPRRSILTRSSSVSPKSSLYLRDSLSHGEKTKQESL 515
 624 -----ASSGPNOCVICLRVSLCPRALRLHYGQHGGGERPKCKVCGRAVSTRGNLRAHF 676

Db 516 LSLQHPSSSTHP---VPLRKSMSWPSA-----ACTISTHHTHTFGSY--SP 556
 Qy 677 VGHKSPARAQNSCPLQCKKFTNAVTLQCHVM---HLGGIIPNGTLLPREGGAQEN 733
 Db 557 DDHVADPEVPSRNT-PV-----FTS-----HPRMKRIHAILEPLGG----- 592
 Qy 734 GSEQSTVSGASFPQOOSQOSP--EEELSEEEDEEEDVDVTD-----ED 779
 Db 593 -----EYSSEBGPSSKDPTSKPSDPEPEKESDLTKTKKGFKTKGANYEC 638
 Qy 780 SLARG------SESGEKAIVRGDSEASGAEEVGYVAAATAGK--E 822
 Db 639 TICGARYKKRDNEYAHKKYCSLEIQTAKHSGVAHEVEKTOAPEPESQMMHKLGAITLE 698
 Qy 823 MDSNEKTQOSSLPPRPPPSLDQPPWEQSSSVYLGGKEBGGKPRSSPASALPEGE 882
 Db 699 LTPLRKRKRKSLGDEEPPAFACPGSETHHNPRLGSTK---SPAEASASPSL-EDPR 754
 Qy 883 ATSVTLVEELSLQAMKREPESSSRKACEVCGAFAFSSQALAEHQTHPKEGPLFTCVF 942
 Db 755 ASSPGLPSQ-----ELGQNGGRG-EGCPKFTVIQHTSSEFKSDPEQP----- 798
 Qy 943 CRQGFLEATLKKHMLAHNOVPPAPRG-----PONTAALSLVPGCSPTS 990
 Db 799 -----SGLEEDKPPAOFSSPPAPRGSAHSIQLPRLVROPNIQVPELIVTEEDPRD 850
 Qy 991 TGLSPFRKDDPT 1003
 Db 851 TEDEPPKEPEKT 863
 RESULT 9
 B60191
 Transcription regulatory protein Evi-1, short form - human
 N:Alternate names: ecotropic viral integration site 1; oncogene Evi-1
 C:Species: Homo sapiens (man)
 C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 29-Aug-1997
 R:Morishita, K.; Parganas, E.; Douglass, E.C.; Ihle, J.N.
 Oncogene 5, 963-971, 1990
 A:Title: Unique expression of the human Evi-1 gene in an endometrial carcinoma cell 1
 A:Reference number: A60191; MUID:90326419; PMID:2115646
 A:Accession: B60191
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-727 <MOR>
 C:Genetics:
 A:Gene: GDB:EV11
 A:Cross-references: GDB:119889; OMIM:165215
 A:Map position: 3q26-3q26
 C:Keywords: alternative splicing; DNA binding; zinc finger
 Query Match 6.3%; Score 330.5; DB 2; Length 727;
 Best Local Similarity 24.2%; Pred. No. 8.9e-09;
 Matches 150; Conservative 82; Mismatches 231; Indels 157; Gaps 30;
 368 GGRHKRCFAVFGSDSALOIHLS-HTGERPYKCNVGNRFETTRGNLKVHHRHREKYP 426
 Db 128 GKHYECENCAVFTDPSNLQRIHRSQHVCAKACBECGTFATSSGLKQKHHSVKP 187
 Qy 427 HVQMPN--HVPEN-----LDVYITSSG---LPYGMVSP-----EKAEEAATPGGG 469
 Db 188 FIFRSQSMYPPDRDLRLSLPLKMEPQSPGVKTKLQGSSESPPDLTKRKDEKPLTP--- 244
 Qy 470 VERKPLVASTALSTESTILLSTAGTATAPDLPAFNKVKVMKAVRP-----KKKA 521
 Db 245 VPSKRPVVPATSDQDPLDLSMGSRSRASGT-----KLTERKKNHVGKKGKS 291
 Qy 522 DENTPPGSEGAISGVAESSTAT-----LMQSLKMTSLSPSMALLTNHFK-STG----- 569
 Db 292 NVESRPAADGS-----LQIARPTPFMDPIYRVREKRLTLQPLALELKKKYLAPSGFLFHP 346
 Qy 570 SFPLPLCARLGAASPESTSKLOQLVEKIDROGAV-----AVTSAAS-GAPTTSAPA 619

Query Match	Best Local Similarity	Matches	Score	DB	Length	2232:
231:	Conservative	128:	Mismatches	435:	Indels	295:
231:	Conservative	128:	Mismatches	435:	Indels	295:
51	ONACSTDPVWYIIGGOENPNSS	-----ASSEDPREGHNPNQVMDTEHSNPDGSSV	104			
498	OSASSSPETMSTVSG	-----PTGSTVTVVPGSSTSPAPSSPMPSS	548			
105	PTDPTWGPERRGES	-----SGHFLVAATGTAAAGGGGLIASPRLGAP	153			
549	TIS	-----GSSITIVSTVSGSTVSGSTGTSQS	593			
154	STPAPPPPPPP	-----PPPGVSGHNLPILEELRVLQORQHOMQ	196			
594	SSPQSSQSPAPMTGSTRPQTSQSPSPMNPSS	-----STPTGSSQSTRPREGSTVASSPTGS	652			
197	-----MEOICRQVLLG	-----SLG-QTVGAPASPSL	232			
653	TGSTRSVAEVTSGSTVSPSSSLCTQSTNSSPSPSLSPSTGSMSTLTSEPPSSQSSG	712				
233	TKPLPLFPPIKPVQTSKTLASSSSSSSSSGAETPKQAFHLHLPLHSQHPSPAGVGR	292				
713	AOSTLTTPSP	-----NSQSTSSLESSTSGATSSGASATYH	760			
293	SHKPPAPSPALPGSTDQLIASPHLAPPTTGLAOCIGAARGLEATSPG	350				
761	SQGST	-----SPAASTTSGEMTSQGSTQTPGSSVSTSAAILTSTQOVSSTNSPGSTVTRPST	817			
351	GGGELSYGE	-----VWGPLEKPGGRHKRCFCAKVFGPSALQIHLRSHNGERYKKNVGNRF	408			
818	VSGSTSSGSTVTVGSTD	-----ASTSSGSA	843			
409	TTKGNLKVHFHRRREKYPHYQMNPNPVPENHLDYVTTSSGLPYGMSVPEKAEBAATPGG	468				
844	-----SSSPAPSTSNQNPNSPSSGSSMITQSPYPQGSTP	887				
469	GYERKPLVASTYALSLATESLTLSTLSAGTATAPAGLPAPFKFVLMAKAVEPKNKADENPPG	528				
888	GSPGTTL	-----TSTSPSPQSTTTIGSTQG	935			
529	SEGSAL	-----SGVNES	581			
936	STGISTVTPSTVDSSTGSSIVTVGSTGSSSPRISTQNTNPSTSSGS	984				
582	ASDSEMSKLOOLEVEKIDROGAVAVTSAASGAP	637				
985	SMSTQTPQSSQSTSPVSSSTSGATSS	1042				
638	LSCPRALRLHYGHGHRPFCKCKVCGRAFS	683				
1043	STISQSTSTPSTGSTVTVKPTSVSGSASSGSTATMGSTEASSTSGGSSSTPNPDSQSTSP	1102				
684	AARQNSCPTCOKKFTTNAVTLQOHVRLHGG	719				

Db 1103 STSGATSSPGSSGTTLTSTISPSPOSSTIGSSOGSTSPVVTSGDMTSOGSTIQIGSTG 1162
 Qy 720 GTALPEGGAADENGSEOSTVSGAGSFPOOQSQOQSPSEFELISEEEDDEEDVEDVD 779
 Db 1163 STVTPSTGSGSTSTGELTISQSTQTP-RSSLSTSPAISTSTQOVSSTNPGCVTQPS 1221
 Qy 780 SLAGGSESGEKGKAISVRGDSEASGAEEVGVAAATAAGKENDSNKETTQOSSLPPPP 839
 Db 1222 TV-RGSTRSSG-----SIVTGSTEGSSSTGSSASATSLSS--SSP 1257
 Qy 840 PPDSILDQOPMEQSGGVLGKEEGKPERSSSPASALTPGEATSVTLVELSLQEAHR 899
 Db 1258 VPSTOSQSPMSTSGSS-----TPTPMPSQSTSPVVS-TTGTGMTSHGSTQTPSTIGSTV 1310
 Qy 900 KRP-----GRSSSRKACEV-----CGQAF--PSQAALAEHOKHTHKEGPLETVCFRQ 945
 Db 1311 TOPSTVSGSSNSGSIYTTIGSSEASTSGSSFKTSPSSIS-----PVPISSPIPTTFASS 1364
 Qy 946 GFLERATLKKHMLAHNOVQFPAHGPONIALSLVPGCSSESTST-----GLSPPEP 998
 Db 1365 --TSGTISIDVSSVSTTSLAPLSSSLP-----STVPSSTQSFSTSGSSKASSSPVPS 1416
 Qy 999 K--DDPTIP 1005
 Db 1417 QTSSTPTNP 1425
 RESULT 12
 T13893
 gene hindsight protein - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
 C:Accession: T13893
 R:YIP, M.L.R.; Lamka, M.L.; Lipsnitz, H.D.
 A:Title: Control of germ-band retraction in Drosophila by the zinc-finger protein HIN
 A:Feature: Control of germ-band retraction in Drosophila by the zinc-finger protein HIN
 A:Reference number: Z17807; MUID:97330681; PMID:9167140
 A:Accession: T13893
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1920 <YIP>
 A:Cross-references: EMBL:086010; NID:g2769709; PID:g2769710; PIDN:AAB95640.1
 C:Genetics:
 A:Gene: hindsight
 A:Cross-references: FlyBase:FBgn0003053
 C:Function:
 A:Description: probably function as a transcription factor
 C:Keywords: nucleus; zinc finger
 Query Match 5.9%; Score 313; DB 2; Length 1920;
 Best Local Similarity 20.0%; Pred. No. 1.7e-07;
 Matches 249; Conservative 137; Mismatches 438; Indels 424; Gaps 56;
 Db 16 GEPAEIGDASEDDHPQVCAKCAQFTDPTFLAHONACSTDPVYVILIGQENPNNSS- 74
 Qy 382 GNTLSAGAESEYD-----ASCS---TDVSSGSHSRSSSS-----LNNNNNNGH 423
 Db 75 -----ASSEPRPEGHNNPOVMDTEHSNPPDSSGVLPDPTWGPERRGEGSSG 121
 Qy 424 KANNNLKLELESTEDDOTENKQRRKRTINNNTIIESEQDEMD-----DEADADAV 478
 Db 122 HPLVATGTGAAGGGGLASPKLGATPLRPESPAPRPPPPPP-----PVGSG--- 173
 Qy 479 AMLTSPDVAT-----LLAGASAGA-----ASASPIPPSPASRALLLSCRACGASDE 527
 Db 174 -----HLNIPLELELRYL-----OQRIHQOMTEQICR-QVLLLSIGQTVG 216
 Qy 528 TPLALCVHLDAMSDIPAKCDCEVIFATHROLQSH-----CCRLPALAGGLPPLLG 580
 Db 217 APASP--SELPGCTASTATKPLPLSPKIKPVQTSKTLAASSSSSSSSSGAETPKQAFH 274
 Qy 581 ASSSPPLNEPDEEHDDDEL-----EOKERLASQSED----- 614

[illegible]

Submitted to the EMBL Data Library, October 1998
A:Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.
A:Reference number: Z17692
A:Accession: T13594
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1891 <FEER>
A:Cross-references: EMBL:AL031227; NID:e1330103; PID:e1316856; PIDN:CAA20227.1
C:Genetics:
A:Gene: peb
A:Cross-references: FlyBase:FBgn0003053
A:Introns: 289/3

Query Match	5.8%	Score 308	DB 2	Length 1891
Best Local Similarity	20.3%	Pred. No. 3e-07		
Matches 247	Conservative 130	Mismatches 425	Indels 416	Gaps 56
QY 53	ACSTDPEVYVIGGEQENPNSSASSEPRDEGIN-NDQVMDTEHSNPPDSSSVPTDPWG	111		
Db 397	SCSTD-----VSSGSHSRSSSLNNNNNNNNKANNNTLDLE-----ELEVSTDDQDEN	446		
QY 112	PERGESSGHFLV-----AATGAAGGGGLILASPRIGATPLRPES	154		
Db 447	KORRLFTTTNNNTIESEQEDMDEADADADAVAMLTSTPDVATLLAGASASGA-----A	500		
QY 155	TPAPRPPPPPP-----PGVSG-----HMLPLLEELRYL-----QQR	190		
Db 501	SKSPTPSPASPLLLSCSPACGASDFETLLPALCVHLDAMHSDIPAKCRDCEVLATRRQL	560		
QY 191	QIHOMQTEOICR-OVLLLSLGOTVGAPASD--SELPGTGAASRKLPLRLPSPIKPVQ	247		
Db 561	QSH-----CCRLPNLAGGLPPLLGASSPLHNEEPEDEHGDDEDL-----E	603		
QY 248	TSKTLAASSSSSSSSSGAEPLPKQAFPHLYHPLGSHSPSAGVGSRHKTRP-----	298		
Db 604	QERLASQSED-----FEHOLYLKHTKAMGCGAISHPSPIKHEPADTK	647		
QY 299	-----APSPALPGSTDLLIASPHILAFPSPTGTLIAOCLGAARLEATASGFL-	345		
Db 648	DLADIGSTLNMSTSSSSFFLNFEQSVNTPNSQSYLSDG-----RQEEBDAQAFT	697		
QY 346	-----LKPNNGSEBLSTGEYWGPLEKPGCGRIKCRCAKYFGS	382		
Db 698	SEFRMKLGEFFCKLCTAVFPMLRALKGNHRVHIGAV-----GPAQPRCNCQPYAVCD	752		
QY 383	DSALQIHLSRSHTEPERPKYCWNGRFTTRGNLKYHF-HRH-----RAK-----	424		
Db 753	KAAIVRHMTTHNDREYECACVANCYATFTTKANCERILRRNKGHTSREEVKRAIYYHRAEDA	812		
QY 425	-----YPIVQNMHPHP-----EHLIDYVITSSGLPYGMSVPEKA	459		
Db 813	GCEDSKSRLEGDLADTYSFRSLISPTPRPPVNESKQLKHM--LLGNNHLIARVNAQOPPLKI	870		
QY 460	EEEAATPGGVVERKPLY-----ASTTLLS-ATESLTILSTASGATAFGPAF	506		
Db 871	QYKSLDQ--LVDRKFPAPARQOQOQOQOQEKGAALDFSDVLDLKKPIGGAS-----	921		
QY 507	NKFVLMKAVEPKNKADENTPPGSEGA-ISGVASSTATLMO-----LSKIMTSL	555		
Db 922	-----LTPAATPTPAALAVPYPPGVGTPTDLAAIEQOQLLAAOQLLAGAGGEYMQQLFRSL	977		
QY 556	PSWALLTNHFKS-TGSEFP-LPLARALGASPSSESKLDOLYVKIDIRQGAVVAVISASGAP	613		
Db 978	-----MFQDSOTSGFPPFPFMA--PPPOANDEKPRMVSPPKNINPMV-GVGVGP	1025		
QY 614	TTSADAPSSASSGPPMOCVICLRYLSCPRALRLHYGQHGGERPFCKYCGRAFTSGNLR	673		
Db 1026	V-----PPGGPVKMYIKNGVL-MPKQQRQRYRT--ERRPAFCEHCSARFTLSNME	1072		
QY 674	AHF-----VGH-----KASPARA-----	687		
Db 1073	RHYVQKHQDFYARQORSHGHVNRGASVNAAAAAAAAAAATYVAGPGSGSGFSNHH	1132		


```

0Y 544 TLMQSLKMTSLPSPMALLTFHNFSTGSEPLD--CARL-----GASPSPT 587
Db 274 -----QLVTHHRRHTGKPRGCSCKRAFFEKSELIRHOTHTGKEPYEC 318
2Y 588 SKLOQLVEKIDROGAAVAVTSASGAPTTAPAPSSASSGPNOCVLCRLVSCPRALH 647
Db 319 SECRKAFR--EKSSLINHOTHG-----EKPHGCIQCKAFKFCOKSHLISH 362
2Y 648 YGOGHGERPFCKVCGRASFTGRLNRAHFYGHKASPARAQNOSCIOCKKFTNATLQOH 707
Db 363 QMTHGKERFKTSCCKAFSRKSQL--VRHOTHTGKEKYECSCECKAPASEKISLTHH 418
2Y 708 VRMLHGGQIPNGTALPEGGGAOENGSEOSTVSAGSFPQOOSQOPSPREBELSEEEBE 767
Db 419 QRH-----TGEKPYVCSCECKAFKFCOKSHLISHORTHTGKEPYE 457
0Y 768 DEEDEDVTDDEDSLAGRGSBGEKKAISVAGDSSEAGAEVEGTVAALAAAGKEMDSNE 827
Db 458 CSECKARKEKSLAHTORTHTGKEKYECH-DECKAFRSQKSLNT-----HQHTHGE 509
2Y 828 KTTQOSSLPPPPPSLDPQPEMGSSGVLGKEBEGKPERSSSPASALTPEGEATSVT 887
Db 510 K-----PYECLCKRAFFEKSELIRLHTHTGKEKPEYECNECKRAFR----- 550
2Y 888 LVEELSLQEMARKPEGESSRKACEVGOGAAPPQOALEEHOKTHPKREPLTTCVFCROGF 947
Db 551 --EKSSLINHQAHTHTGKEKPE--CSECKARFASRKSHPILHORTHTGKEKPYGCSCECKRAF 605
2Y 948 LERATLKHHMLLAHHOVPF 967
Db 606 SOKSOLVNHORI-HTGKEKPY 624

```

```

accession | 0
resolution | 0.00647
finger | protein - African clawed frog
species | Xenopus laevis (African clawed frog)
date | 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 31-Dec-1993
accession | S00647
author | Ruiz I Altaba, A.; Perry-O'Keefe, H.; Melton, D.A.
journal | Mbo J. 6, 3065-3070, 1987
title | Xfin: an embryonic gene encoding a multifingered protein in Xenopus.
reference | number: S00647; MUID:88082679; PMID:2826129
accession | S00647

```

```

;Molecule type: mRNA
;Residues: 1-1350 <RUI>
;Cross-references: EMBL: X06021
;Note: it is uncertain whether Met-1, Met-11, Met-16, Met-38, or Met-39 is the initiator
;Genetics:
;Gene: fin
;Keywords: DNA binding; zinc finger

```

Query Match	5.6%;	Score 296.5;	DB 2;	Length 1350;
Best Local Similarity	21.3%;	Pred. No. 7.3e-07;		
Matches 131;	Conservative 64;	Mismatches 249;	Indels 171;	Gaps 177;

iy	365	EKGGGRHKRCRCAPVFGSDALQILHNSHTGERPKCVNGCRFTTNGNLKVFHRRHK	424
	111	-----	
ib	407	EKP---FKSCHCDKFKFTERRSLAHORHTTEKRPKCDCKEKFORSNLILHORITGE	463
	111	-----	
iy	425	YPHVQMNPHRVEPHLDVITVSSGLPYGMSVPRKEAEELAPGGGVKKPLVASTTALSA	484
	111	-----	
ib	464	RPV-----	466
	111	-----	
iy	485	TESLTLTSTAGTATADGLPAFNKFVLMKAVERPKKADENTPPGSEGSALGVAAESTA-	543
	111	-----	
ib	467	--KCTLCDRT-----FIONSDLVKKRQKVANLP-----LSDPHYAN	500
	111	-----	
	544	TLMLQSLMTSLPSMALLTLNHFKEGTSGSPRLPCRALGASPSSTKLQQLVEKIDRQAV	603
iy	501	SPFKGCKCDLTFBSHWTFPMKSHKIHSEKKFOCKAEC-----KKGFYOKSDLVNHI	550
	111	-----	
iy	604	AVTSAAGAPTTTAPADSSSASGPNQVCILRVLSCPRAIRLRLHYOHGGERPFKCKYCG	663

```

Db      551 RHVHTGEK-----PFXCOLLCKSEFSONSDLKHWRHHTGEKPFPCYTOD      593
OY      664 RAFTSRGNLRAHFVGNHKASPAARAQNSOPIQCKFTTNVTLQOJHRMHLGGQIPNGTAL      723
Db      594 KSTERSALIKH---HRTHTGER-PHKSVCQKGFIOKSAIHKSRHT-----        637
OY      724 PBGGGAAOENGSEOSTVSGAGSPFOQSOQSPREBELSEEEDEEEDVTDJEDSLAG      783
Db      638 -----TGERPYTOQCGKSTIQNSDLVKHQRHHTGEKPYHTECKNRTTRESSLYK      688
OY      784 KSESGGEKAIIVRGDSEASGAEEV-----GYVAAATAGKEMDSNEKTTOOSSLP      836
Db      689 HRTHSGEKPYPCQCEKFTIQSSDLVKHLVHNGENPAPATAFHEILLIRRENLIRSEED      748
OY      837 PPPPPSLDOPQMEQSSGVLGKEGEGKPERSSPASAALT---PCEATSVLVEEL      892
Db      749 RYPCYTC-----GKVPNRPALLKHLRTHTEKRYPCNECDK-SFQFS      791
OY      893 SIQEARKEGESSSKRACVCOAFPSQALAEHQKTHPREGPLYTFVFCQGGLEART      952
Db      792 DLYVKNLRTHTGERPYH-CECNKGFIONSDLVKNRHTHTGERP-YTCSQCDKGEIORS      848
OY      953 LKKNMLAHNOYOF 967
Db      849 LTKNM-RTHTGEKPY 862

```

RESULT 17
T46277
hypothetical protein DKFZp564M2178.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46277
R:Ottensmøller, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23031
A:Accession: T46277
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-64 <AA>
A:Cross-references: EMBL:AL137516
A:Experimental source: fetal brain; clone DKFZp564M2178
C:Genetics:
A:Note: DKFZp564M2178.1

Query Match Similarity	5.6%;	Score	296;	DB	2;	Length	644;
Best Local Similarity	24.6%;	Pred. No.	3.5e-07;				
Matches	179;	Conservative	70;	Mismatches	282;	Indels	198;
				Gaps			35;
QY	299	ADSPA-LPGS-TDQIASPHLAFSTTGLLAQAQCGAANGLEATASPGILKPRNGSGE--	354				
Db	3	ASSPAEVPVSQPPDLPAASDH-----SYELRNGEALIGRDR-----RGRARRNNNSEAG	50				
QY	355	-----LSTGEVWGPLEK-PCGRHKRCRCAKYFGSDSAQIHLRSHTGER	397				
Db	51	GAATQELFCSACDQLFLSPHQLOOHLSHREGVEKFCPLCSRVPFSPSSLDQHLGDHSES	110				
QY	398	PYKCNVCGMRFTTRGNLKVHNRHREKRYPHQOMN-HPVEHLIDY-IIS-----	445				
Db	111	HFLLDCDCAFGAFTGELLALHRRATP-----NFLHSCPCGKTFVNLKFLYHRRITGV	163				
QY	446	SGLPYGM-S-VPEEKAEEEAATPCGGVERKPLVASTTALATSLTLTSSAGTATAPG--	502				
Db	164	GGVPLPTTPVPE-EPVIGFP-----EPAPAEITGEDEAPRPVPSSEETSAAGA-APGTY	214				
QY	503	-----LPAFNKFLVLMKAVEPKKKADENPPTSGEGSAISGYAESSTATIMOL	548				
Db	215	RLCLCSREFGKALQLLRHQRFV-HRLERRRKC-----SICG-----KM	251				
QY	549	SKLMTSLSDSWALLTHFKS-TGS-FPLPLCARALGASPSSTSKLQQLVEKIDROGAVAV	605				
Db	252	FKKSH-----VRNLLRHTTGGERPFCPCPSCKPFN-SPANLAR-HRLNHTGGRPYRGCD	303				

[illegible]

RESULT 18
S70619
finger protein bowel - fruit fly (*Drosophila melanogaster*)

C:Accession: S70619
C:Date: 19-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 17-Nov-2000
C:Species: *Microphid mehinogaster*

R:Maeng, L.; Coulter, D.E.
EMBO J. 15, 3182-3196, 1996
A:Title: bowen, an odd-skipped homolog, functions in the terminal pathway during Drosophila embryogenesis
A:Reference number: S70619; WUID:96272178; PMID:8670819
A:Accession: S70619

A:Accession: U774 NM_017447
 A:Cross-References: EMBL:U58282; NID:g1388165; PIDN:AA817949.1; PID:g1388166
 C:Genetics:
 A:Gene: bow1
 A:Cross-References: FlyBase:FBgn0004893
 A:Map position: 2L

Query Match	5.6%	Score 293.5;	DB 2;	Length 744;
Best Local Similarity	17.2%	Pred. NO. 5.4e-07;		
Matches 174; Conservative	88;	Mismatches 327;	Indels 425;	Gaps 24;

QY 90 MDTEHSNRPDSSGSSPTDPTMGPERRCEESSGHFLVAATGTAAGGGGILLASPRLG--- 146

Db 1 MPTSSSSSEISGGGGCAIPMLRPSRMDQFMS-----MAAAAAAVGGGILIPGAADRNGSG 56

0y	----	ATPPESTP-----	APPPPPP-----	164
147	----	ATPPESTP-----	APPPPPP-----	164
		1 : : : 1	1 1 1 1 1 1	
57	CCCCCCCCCCCCCCCC	CAVETOTAVOOURACTICUPPCCCCCCCC	CCCCCCCCCCCCCCCC	116

```

165 --pppgvsgshlnlplllelrvlqoqiohqomnteqicrvlllgsllgqtvgapasps 222
      || | : | : : : | | : || : | | |

```

223 ELPGGTGTA~~STK~~PL~~LP~~LFSP~~IK~~VPQTSK~~L~~IA~~SSSSSSSSSS~~SGAETPK~~QAF~~HL~~YH~~PLGSQ 282

Db 165 -VPPPGV-----LYGPAGVP 178

Db 179 PDPYLGT-----PGSPPTGAGS-----SP---FPPGAAALFPFGCLPGCMHAGLD 2211

```

343 PGLLKPNAGSGEISGEVWCPLEFGGHTNCFANVFQSDSRQILNLSHOLN INCR 402
      ||: | :| :| :| | | | | | :
222 RRLR-----APGRASRPKQFICKFCNRQFTKSINLLIHENTHTDERPYSCD 269

```

QY	403 VCGNRFETTRGNLAKENHNRHKRYPHVQGNMPPRVENDLDVIITSSGLPYGMSVPREKAEEL	462
Db	270 ICGKAFFRQDHLRDHRYINSTRK-----	293

Qy	463	AATPGGVERKPLVASTTALSATESLTLTSTAGIATAPGLPAFNKFLMKAVEPNKAD	5222
Qy	294	-----	293

QY 523 ENTPPGSGSAISGVAESSTATLMQLSKMTSLPSSWALLTNHFKSTGSEPLPLCARALGA 5822

QY 583 SPSETSKILOLVEKIDROGAVAVTSAASCAPTSAAPSSSASGPNOCVICLRVLSCPR 642

DZ
643 ALRLHYGGGERPEKVCGRPFSTRGNLRHFVGHKASPARAQNCSPIQKKFTNAV 703

OY

Db 308 TLAVHKILMEESPCHKVCSSFSNFQSRNKTLLTH---TDHKPYECSGCGKVFRRNC 36

```

Db      364  DIRKHALTAVGEVNSGDYVDVGEEDEARLISGDEEDSLLEVDSRQSPHNLGESSGS 42
      731  OENGSEFOSVSGAGSFPPDOOSOP--SPPELSE-----EEEEDEEEEDVYDE 77

```

```
Db      424 EKSESERMRLKRAALDHEESEEFDDFDEEEELQDLPVHDLPREDDDDDFPEDEEOAE 48
```

Db 484 VALVARQASKAATSOSSSSVGTKEPERQGYTHCHHECGEYTMRPHPGHEKHDEEPGNGSI 54

```

053 S$FFFF-----FFUSUUUFFEHEGGSSVV DQAKUDDGK LNDGDKT   GHTA  19
      :||| || | : | | | | | :
Db  544 ASLVPVPSFVRYSVPPGAAGPPP-----APPGARPPTHQHPRGHDLPLRPNGD-  59

```

884 TSVTLVEEISLQEAMRREPESRRKACEYCGAIFPSQAALLEHOKTHPKPELFTCYVC 94
 591 -----PYLEPLIHVRDLHNKSLNLSKAGVPPRPHTPTITTPESGKPLANPLHSPNEA 64

QY 944 RQGFLEATLKKMLAHNQVOPARPHGQPNLAALSLVPGCSPTSTGSLSPR 997
 945 MPESTGSIAPMKRTPL-PTIDIMDPNHNPGSGORTFYD--SPSYALNMRNP 695

RESULT 19

zinc finger protein ZNF91 - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 05-Nov-19

C/Accession: 552305
R/Bellefroid, E.J.; Marine, J.C.; Ried, T.; Lecocq, P.J.; Riviere, M.; Am
EMBO J. 12, 1363-1374, 1993

A:Reference number: S35305; MUID:93223677; PMID:8467795
A:Accession: S35305
A:Molecule type:mrna

A; Cross-references: EMBL:L11672; NID:g186773; PIDN:AAAS9469.1; PID:g18677
A; Note: the authors translated the codon GCA for residue 750 as Thr and G
C; Genetics:

A:Gene: GDB:ZNF91; HPE//; HNF10
A:Cross-references: GDB:132284
A:Map position: 19p12-19p12

Query Match	5.68;	Score 293.5;	DB 2;	Length 1191;
Best Local Similarity	22.5%;	Pred. No. 8.9e-07;		

```

Matches 141; Conservative 75; Mismatches 239; Indels 173; Gaps 24;

|OY 365 EKGGRRHRCRCACVFGSDSALOIHLSHTGGERPYKCVNCGNRFRTGRNLKVHFNHREK 424
| 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|Db 627 EKR--YKCECGCAFAHSSSALAKHNRHITGGEKPYKKECGKAFNSNSTLANHKHKTHTEE 683
|OY 425 YPIVQAMPHPYREHLDIVYITSSGCLPYKGVPRPEKAEEDAAATGGGVERRPLVASTTALA 484
| 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|Db 684 KPRKCKECBDTKFKRLSTLTNKHIIHAGEKL--YKCEE-----CGKAFNR 725
|OY 485 TETLTSTLSAGTATAPLPAKNEFVLMKAVEPKNKADENPTPGESEGAISGVASSTNT 544
| 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|Db 726 SSSLTI-----HKFI--HTGEKPYKCE-----CGKAFNWS 756
|OY 545 LMOLSKIMTSLPS-----W-ALLTNHKK-STGSEPLPL-CARALGASPSETS- 588
| 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|Db 757 LTKHNRHITREKPRKCEKCGKAFIWSSTLTNRHNRHITGEKPYKCECGKAFRSRSLTLTKH 816
|OY 589 KLDQALVEKIDROGAVAVTSAASCAPTTSAAPRSSASSGPNOCYICLRVLSCPRALRLHY 648
| 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|Db 817 KTHITGEK-----PYKCECGKAFKHSSSALAKH 845
|OY 649 GQHGGERPRCKVCYGFSTGKILRAHFVGNKASPARAONSCPTCOKKTFTNAVYLOOHV 708
| 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|Db 846 ITHAGEKLYKCECGKAFNOSNLTNKHIIHTEKPSKSE---CDKAFIWSSTLTLEHK 901
|OY 709 RMHIGQIPNGGTAIPLPGGGAOENGSEDSYVSQASFPQOOSQPRPEELSEEEED 768
| 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|Db 902 RII-----TREKPYKCECGKAFS-----QPSHLTTHKRNHTGKPYK 940
|OY 769 EEEEDVTDEDSLARGSESGEKAISVKGDSSEASGAEVEGTVAATAATGKEMDSNE- 827
| 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|Db 941 EECGKAFSOSTLTNKHIIHTEKPYK---CECGKAFKSSTL---TEHKIITHTEK 992
|OY 828 -----KTQOSSLPRPPRDSLDQPRMEOGSSGVLGKEGCKRPSRSPASALTP 879
| 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|Db 993 PKCECGKAFSSS-----TLTRHTRHNGE-----KPYKCECGKAFNR 1033
|OY 880 EGEANSTVLEELSLQAMRKEPESSSRKACEVCGAFPSQALAEHOKTHPEKPLET 939
| 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|Db 1034 SSKLTHTNKTII-----HTGEKPYK--CECGKAFISSLTNHNKRIHTEKPR-YK 1079
|OY 940 CVFCRQGLERATLTKHMLLAHNOYQRP 967
| 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|Db 1080 CECCGKAFSSOSTLTNKHRL-HTGEKPY 1106

RESULT 20
I43676
hunchback-related protein - Caenorhabditis elegans
2:Species: Caenorhabditis elegans
3:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
4:Accession: T43676
5:Fay, D.S.; Stanley, H.M.; Han, M.; Wood, W.B.
6:rev. Biol. 205, 240-253, 1999
7:title: A Caenorhabditis elegans homologue of hunchback is required for late stages of
8:Reference number: 22622; MID:99117349; PMID:9917360
9:Accession: T43676
10:Status: preliminary; translated from GB/EMBL/DBJ
11:Molecule type: mRNA
12:Residues: 1-982 <FAV>
13:Cross-references: EMBL:AF097737; NID:g4323034; PIDN:AMD16170.1; PID:g4323035
14:Genetics:
15:Gene: hbl-1

Query Match 5.5%; Score 292.5; DB 2; Length 982;
Best Local Similarity 18.6%; Pred. No. 8.1e-07;
Matches 214; Conservative 134; Mismatches 394; Indels 409; Gaps 46;

|Y 5 SERSRLG-----VAGE-PAELG--GDAS-----BEDHQVCAKCAQFTDPT 45
| 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|Db 49 SANNNSLDNMNCCVFGSELPQHIGPFGNSSLTMTTAQDPGKIHDP-----GGIVSPK 102

```

QY	46	E----	FLAHONACSTDPVWVLIIGCGENPNNS\$ASSEPRPECHNNPQVMDTSHNPDSG	101		
Db	103	EDGRKSEHTNSYD-----	VSASGSPNDGAQSDSTSDHIDVEMCTETEMTDDEK	154		
QY	102	SSVPTDTPWBERGEESSGHFLVAATGAAGGGGLLASPKLCATPLPRESTAP---	158			
Db	155	STIKPEQATPKLEEGSDSKPESTSVETS-----	SNQVSEVQMOMPIPIYIP	205		
QY	159	-----	PPPPPPPPPGVSGHINIPILTEELRVLOQRQHOMOMTQICQVOLLISL	211		
Db	206	SFLKNSLPAPILPIPTPGOSANVERNSNSIEALL---	TISQOFAFVFAAARIKRS	261		
QY	212	GQVYGAPASSELPGTSTASTKPLDLPSFKYQVQYKSTLASSSSSSSSGAEPTKQA	271			
Db	262	SESIGORS-----	GTSA-----	FLNIEPEKMSANNMNEEA-----	295	
QY	272	FFHLHPHLSQHFFSAGVGRSHKPTAPSPALPGSTQQLASPIPLAPSTTGLAACOL	331			
Db	296	-----	PASTVSACSTP-----	TTTTSASFC-	315	
QY	332	GAARGLEATSPGLLPKNGSELSTYGVNGLPEKPGGRH-----	KCRFCAKYFGSDSA	385		
Db	316	-----	RPGL-----	GPVALPTQONGOTPMVLCPIGCFMCKSPKH	350	
QY	386	LQHLRSHTEGPERYKCVNCGNRFRTBGNLWVHFRHREKKYHYQMNHPRPVPHDVTYS	445			
Db	351	FNSHMNTH-GD--	HQSCMCTYSTETBERLKNH--	RES-----	HTVEQOLRAGES	396
QY	446	SGLPYMSVPERKEAEBAATPGGGVERKPLVASTTALSTATESTLILSTASATAPGIPA	505			
Db	397	-----	EPAKESASP-----	KNLSL--	SKDGSATSPINEI	424
QY	506	FNFVLVLMKV--EPKKNADENTPPGSSEGSALSVAESSTALTMQSKLMTSPSALLTNH	564			
Db	425	FNLSITMASTLDSFTN--	-----	AVSISTTEQ\$AL-----	SALTLD	460
QY	565	FKSTGFPPLCARALGASPSSETSKLQOLLEKIDROGAVAVTSA--	ASGAPTSAPAPS	621		
Db	461	MSSTPILSTLASHSPGVSALDQIKAI\$ENSPFMEGGINL\$AL\$GVVSNAIKKDTSP\$E	520			
QY	622	SS-----	ASSGP-----	NOC-----	VICLRVLSCPRA-----	643
Db	521	K\$N\$GECRRSS\$SKIKIFPKCQGHOSL\$KDDOMAHARTHI\$PAEKOLNCHONCVTEYKH	580			
QY	644	-LRLNHGQ\$G\$E\$RPFCKCYOG\$RA\$FT\$RGNLRAHFVGH\$K\$AP\$ARA\$ONS\$C\$P\$CK\$KFTNAV	702			
Db	581	HLEHYH\$RNIHG\$K\$P\$Q\$CKK\$CAY\$K\$V\$K\$M\$N\$H\$K\$H-----	TNH\$Q\$R\$C\$M\$D\$C\$Y\$AT\$AT\$K\$H	636		
QY	703	TLQO\$H\$R\$M\$LG\$Q\$IP\$NG\$T\$AL\$P\$R\$G\$G\$A\$Q\$E\$N\$S\$E\$O\$T\$V\$S\$G\$S\$FP\$Q\$O-----	OSQ	752		
Db	637	SLKLHLK\$Y\$N\$H\$R\$V\$P\$G\$--	IEM\$G\$D\$S\$P\$P\$T\$AT\$IT\$P\$S\$P\$IM\$K\$O\$E\$IK\$T\$E\$V\$P\$V\$S\$IA	693		
QY	753	QPS-----	-----	EEEL---	760	
Db	694	Q\$P\$F\$P\$M\$M\$G\$N\$G\$N\$G\$L\$N\$F\$A\$N\$M\$L\$N\$K\$H\$LD\$V\$G\$L\$M\$G\$L\$R\$N\$V\$M\$S\$L\$K\$S\$AC\$D\$F\$V\$A\$S\$AD\$E\$K\$R\$H\$S	753			
QY	761	-----	SEEEEDDEEEDDY--	TDEDSIAG---	RGS	786
Db	754	MSHILN\$S\$N\$V\$P\$T\$IAS\$LY\$N\$L\$M\$P\$S\$H\$V\$A\$P\$D\$M\$N\$A\$E\$M\$D\$C\$V\$K\$ID\$D\$D\$IT\$S\$H\$C\$Y\$E\$M	813			
QY	787	ESGGEK\$AIS\$VR\$D\$SE\$A\$Q\$AE-----	EV\$G\$V\$A\$A\$A\$T\$A\$K\$R\$E\$M\$D\$N-----	EKTQO\$SL	835	
Db	814	D\$G\$S\$N\$A\$V\$P\$T\$G\$S\$O\$T\$S\$D\$E\$E\$F\$K\$K\$C\$K\$S\$L\$E\$O\$T\$S\$R\$A\$M\$O\$N\$N\$S\$P\$M\$N\$D\$S\$A\$M\$E\$K\$D\$G\$E\$AD	873			
QY	836	PPPP\$P\$D\$S\$LD\$Q\$P\$M\$E\$Q\$S\$S\$G\$V\$L\$G\$K\$E\$E\$G\$K\$P\$E\$R\$S\$S\$P\$A\$--	ALTEP--	GEAT\$V\$T\$V\$IE\$E\$L	892	
Db	874	APH\$S\$D\$T\$V\$P\$S\$P-----	PLH\$S\$S\$V\$A\$P\$IP\$IT\$Q\$P\$N\$F\$O\$S\$I\$LA\$Q\$A\$S	916		
QY	893	SLQ\$E\$M\$K\$R\$E\$P\$G\$S\$S\$R\$K\$AC\$E\$V\$C\$Q\$A\$P\$S\$O\$A\$LE\$H\$O\$K\$T\$P\$K\$E\$S\$P\$L\$T\$Y\$C\$F\$C\$Q\$G\$E\$T\$ERAT	952			
Db	917	LIG\$P\$LL\$A\$N\$R\$P-----	S\$A\$F\$Y\$C\$D\$H\$C\$K\$IP\$E\$D\$T\$O\$V\$L\$D\$S\$H\$R\$F\$H\$T\$P\$G\$N\$P--	EM\$C\$D\$C\$Q\$Y\$O\$A\$F\$N\$E\$LS	971	
QY	953	L\$K\$N\$H\$L\$A\$N\$H\$Q	963			

2Y	571	FPLPLCARLGLASPSETS-----KLOOLVEKIDROGAVAVTISAASGAPTTSAAPASSA	624
Db	672	GALPKAAPRSSSSSSSSSSCKSKRYKLOSKVAVLIREGVSSITPA-----KDDSS	719
2Y	625	SSGPQOCVICLRVLSCEPRALRLHYGHHGGERPFCKYCGRAEFTRGNLRAHFV-----	677
Db	720	SSGLGS-----IGVTSRDRRESNPFLLKPDERSP	748
2Y	678	-GHKASPPA-----RAQNSCPICOKKFTNAVTLQOAHVRLHGGQIPNGTALPEGG--	727
Db	749	AEGVAVPAGSTKPKTKAKAKAGAKKAKGKTGKTKPSKTRK---KVRSGGSSTASGGPG	804
2Y	728	-----GAQENGSGOSTVSG-----AGSFPOQOSQOPSP-----	756
Db	805	SUKSKADSCSQAASAKTEETSWSGERITTKAPTPPKVAPPALATPDQSOTVDSCK	864
2Y	757	-----EELSEEEEEDEEEDVPTOEDSLAGRSESGEKAIS	795
Db	865	TPDVSLAEASDITGVYVGAEEEEEEEEEEEOQPATTTATISTAAPATAPSAGS	924
2Y	796	VRGDSSEASGAEEVEGTVAATAAGKEMDSNEKTTQOSSLP PPPPPDSLDQOPMEGSS	855
Db	925	TAGD-----SGAED--GPAARA-----SOLPTLPPMPNNLLPAGVDOCTS	962
2Y	856	GVLG-----GREGGKPERSSS-----PASL--TPREGATSVTLVEEL	892
Db	963	GVYALTLALLFKMEANLASRAKAOELLQDITNOILRHRRPPSTLGVTPAVPVPSFGLPPAP	1022
2Y	893	S-----LOEAMRKEPEGESSR---KACEYCGOAFPSQALAE-	926
Db	1023	SSVLLPGLSLPIGGCGSPPTPTGLVPASDRKREGSSSEGRGDTDKLKLHTEGRAVEEV	1082
2Y	927	-----HOKTHPKBSPLFTCVFCROGFLERAVLKKHMLLA	960
Db	1083	KLISIPYOKKDIITKEEYKDILRKAVHKLCHSKSGEINP---VKVSNLVRAVYQRYRYFR	1139
2Y	961	HHQYOPFARHPR	972
Db	1140	KHGKRPDPDPGR	1151

```

RESULT 26
fzinger protein MZF1 - human
C:Species: Homo sapiens (man)
C:Date: 21-Apr-1992 #sequence_rev: 21-Apr-1992 #text_change 01-Dec-2000
C:Accession: AA0751
R:Homomol. R.; COLLINS, S.J.; HICKSTEIN, D.; RASKIND, W.; DEAVEN, L.L.; O'HARA, P.; HAGER,
U. Biol. Chem. 266, 14183-14187, 1991
A:Title: A retinoid acid-responsive human zinc finger gene, MZF-1, preferentially expres
A:Reference number: AA0751; MUID:91317761; PMID:1860835
A:Accession: AA0751
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1485 <HRO>
A:Cross-references: GB:M58297; NID:g189043; PID:g189044
C:Genetics:
A:Gene: GDB:ZNF42; MZF-1
A:Cross-references: GDB:125898; OMIM:194550
A:Map position: 19q13.2-19q13.4
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 5.3%; Score 277.5; DB 2; Length 485;
Best Local Similarity 21.1%; Pred. No. 2e-06;
Matches 149; Conservative 42; Mismatches 209; Indels 305; Gaps 26;

OY 305 PGSTDOLASPHLAPSTTGL-LAAO-----CL 331
||| |||| |::|
||| |||| |::|
1b 21 VSPHLHVPMDLGMAGLSGQIQSPSREGFAHRYLLPSDLRSEQPTDEDPCR 75
||| |||| |::|
||| |||| |::|
OY 332 GARGLEAF-----ASPKLKPKNSSGELSYGEVMPLEKPGGRHKRFFCAKVVGSDS 384
||| |||| |::|
||| |||| |::|

```

```

Db      76 GVGPAITLTHWRBSPRGSRG--RPSTGGGVVR-----GGR--CDVCGKVTSQRS 120
OY      385 ALOIHLRSHGTGERPYKCVNCGNRPETRGNLKVHEHHRREKYPHVOMNPHVPBHLDYIT 444
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      121 NLIRHQIKIHGERPFVCGSEGRSFRSSHLRLHOLTHTHEER-----FVCG 166
OY      445 SSGLPFGMSYPPPKAEDEEATPGGVERKPLVASTALTATSLLTSTAGATAPGLP 504
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      167 DCGQGF--VRSARLEHHRVHTG--EQPRCAEGGGSFGRSNNLQHORIHGDPPG-- 218
OY      505 AFNKFLMKAVEKRNKADENTPFGSEGAISGVVESTATLMQLSKMTSLPSWALLTNH 564
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      219 -----PGAKPPAPPGAP----- 231
OY      565 FNKSGSFPLPLCARALGASPSYTSKLOOLVERIDROGAVAVTSAAGADTPSADPSSA 624
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      232 --PPGFPPCSECR-----ESFARRAVLLEHQAVH-----TGDK 262
OY      625 SSGPNOCVLOLRVLSCPRALRLHYGHGGERPPCKYCGRASTGNLRAHPGHKAS--P 683
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      263 STG--CVCEGGERFGRRSVLLQHRVHSGERPFRAECGGQSTROKSNLQHRRIHTGERP 319
OY      684 AARAAQSCRICQKFKTNAVYTLQOHVRMHLGGQIPNGTALPREGGAAYDENGSEOSTVSGA 743
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      320 FA-----CAECGKAFRQRPILTLOLRVHTGER-----PRAPCEG-- 354
OY      744 GSPFOOQSOQRPSEBELSEEEDEDEEDVDYDSDLAGRSESGGEKAISVRGDSEEA 803
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      355 -----QFESRLKLTTHORHT-----GKK----- 374
OY      804 SGAESEVGTVAATAATAGKEMDSNKTQOSSLPDPDSDLDOPRMEGSSGVLGKEE 863
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      375 -----PYHCGEGCL----- 383
OY      864 GGPKESSSPASALTPEGEATSVTLVEELSLQEDAMRKDESSSRKACEVCGQAPPSQA 923
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      384 -----GFTQVSRLT--EHORIHGTGERPF--ACEPCGGSFROHAN 418
OY      924 LEBHOKTHPREGFLFTCVFQROGFLERATLKKHMLLANHOVRF 968
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      419 LTHORRIHTGERP--YACPECGKARQRPILTQH--LTHIRREKRF 461

```

RESULT 27
 JC5146
 arylphorin gene-specific binding protein 2 - flesh fly (*Sarcophaga peregrina*)
 N:Alternate names: ABP-2
 C:Species: *Sarcophaga peregrina*
 C:Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 01-Dec-2000
 C:Accession: JC5146; PC4244
 R:Adachi, N.; Kubo, T.; Natori, S.
 J. Biochem. 120, 1239-1246, 1996
 A:Title: Purification, characterization, and cDNA cloning of ABP-2 (arylphorin gene-s
 aga peregrina).
 A:Reference number: JC5146; MID:97164029; PMID:9010776
 A:Accession: JC5146
 A:Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-594 <ADA>
 A:Cross-references: DDBJ:D86231; NID:g1753198; PIDN:BAAL3048.1; PID:g1753199
 A:Accession: PC4244
 A:Status: Preliminary
 A:Molecule type: protein
 A:Residues: 150-157/193-204 <AD2>
 C:Comment: This protein is a negative transcription factor of the arylphorin gene.
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
 F:246-251/Region: alanine-rich
 F:258-271/Region: glutamine-rich
 F:401-411/Region: glycine-rich
 F:413-420/Region: histidine-rich
 F:480-496/Region: asparagine-rich

IY 364 LEKPGGRHKRCFAKAVKVGSPSALQILHLSHTGGERPKYKVCNCNRTTRCNLKHFNHRE 423
 Db 511 TPSPFGSS-----GSHSMSPAVADONHPIISR-----ASNSSSGGPM-----Q 548
 IY 424 KYPHVOMNPHEHDLYVTSSGLPYGMSVPPPEKAEEBAATPGGVERRKFLVASTTALS 483.
 Db 549 GPVPVGAAGPPMPPH-----PGMPGG-----PPGQQQSQQQQ-----ASNASS 587
 IY 484 ATESLTLTSTASATATAPGLPAFNKKFVLMKAVERKNKADEN---TPPSGEGSA----- 533
 Db 568 ASNS-----PQQTPPPAP-----PPNQGNNNATPPPPQGAAGGYPMP 627
 IY 534 -----ISGVAESSSTATLMQSLKMTSLPSMALLTNHFKTGSPP-----LPLCARALG 581
 Db 628 PHMHGCKKMGCPGQSPGAPQYPPPOQPOQYP-----GATPPRPPQYPPGAYATG 675
 IY 582 ASPSETSKLQQLVEKIDRQGAVAVTSAASGAPPTTSAFA---PSSASSGPNOCVICLRVL 638
 Db 676 PPPPPTS-----QAQAGCANSMPSGAQAAGGYPGKGMNHTHQYPPYQWVPPSPQ 725
 IY 639 SCPPRALRLHYGQGGGERPFCVKYGRAFSTRGNLRAHFYGHKASP---AARQNSCPIC 694
 Db 726 TVPG-----GAPCG-----AMGNHYQKGTPTPPPVGGGPPPPGSGSPRLN----- 768
 IY 695 OKFTNAVTLQOHVHRMLGGOIPNGTALPEGGGAOENGSEOSTVSGASFPQOOSQP 754
 Db 769 -----YLKQHLQ-HKGGY---GGSPTPPQPGQGYGNGP-----TGMHGMMPGPP 809
 IY 755 -----SPEELSEEEDEDEEEEDVTDSDSLAGRSSESGEKAISY 796
 Db 810 HHMGPHGFPNNMCPPTSTPPQOSMLQGGQPPQG-----ASGGPESGGPEHIS 857
 IY 797 RGDSEERASGAEEVVG---TVAAATAAGKEMDNKTKTOOSLTPP----- 838
 Db 858 QDNGISSGPTGAAGMAHYAVSVYTTGPDGTSMDYSSQGSTLSNMAASGEDPQCTTPKSR 917
 IY 839 -----PPPDSLDQPPPMEOGSSGVLGKKEGGKPPENSSSPASALTPEG 881
 Db 918 KNDPYQSGLHAPSTSPHPVVMHPG-----GGP---GGEYDMSSPMPMPAG 962
 RESULT 33
 A49073
 HSN motor neurons migration regulator (alternatively spliced) Egl-43 - Caenorhabditis el
 C/Species: Caenorhabditis elegans
 C/Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 01-Dec-2000
 C/Accession: A49073; B49073; T24250; T24249
 R/Garriga, G.; Guenther, C.; Horvitz, H.R.
 Genes Dev. 7, 2097-2109, 1993
 A/Title: Migrations of the Caenorhabditis elegans HSNs are regulated by egl-43, a gene e
 A/Reference number: A49073; MUID:94040732; PMID:8228480
 A/Accession: A49073
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-581 <GAR>
 A/Cross-references: GB:S66757; NID:9452950; PID:9452951
 A/Note: sequence extracted from NCBI backbone (NCBIN:139438, NCBIF:139439)
 A/Accession: B49073
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 222-581 <GA2>
 A/Cross-references: GB:S66936; NID:9452952; PID:9452953
 A/Note: sequence extracted from NCBI backbone (NCBIN:139899, NCBIF:139727)
 R/Wilkinson, J.
 submitted to the EMBL Data Library, October 1995
 A/Reference number: Z19863
 A/Accession: T24250
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-537, 'VPP', '542', 'V', '<MID>
 A/Cross-references: EMBL:666515; PIDN:CAA91353.1; GSPDB:GN00020; CESP:R53.3a

Db 338 PEDFLKASLAVEMERPTYL-----THSPLPSSTT--PSPPASSPEQSLKSSSPHSSP--- 388
OY 101 GSSVPTDPDPWGERBERREESSGHFLVATGTAAAGGGGLILASPKGATPLPESTAPPP 160
Db 389 GNTVSPDLAPGLDPHRDSYS---YLVNSYG-----SEGIGSY---PGYADAPHL 430
OY 161 PPP-----PPPGVSGHILNIPILILELAVLOQRQIHQMOMTEOICROVLL 208
Db 431 PPAFTPSYNAHPKFLPLPGISSNGLS---TMNNINININNSL--FRLPYVYNSMLSG 485
OY 209 GSLGDTGAPAS-PSELPGTGTAASSTKPLPLPSPDKPVQTSKTLAASSSSSSSSSSGAET 267
Db 486 SSLPRLMPLPASPLSSLPTDG---ARRLLP-----PEH 515
OY 268 PQQAFPHLYHPLGSOHPFSAGVGRSHKPPAPSPALPGSTQDLASPLAP-STITGL 326
Db 516 PREVLIRAPH-----SAFSLTGAASMKDESSPPSGSP-TAGTAATSEHVQKATSSVM 569
OY 327 AAQCLGAAAGLEATASFGLLKPNGSELASYEVMGPLEKPGG--RHKCFCAKAVGSDS 384
Db 570 AARPTDGAANL-----IKKRMNTGYKTLPLPLKQMGKIKYEONCAKFTGOLS 619
OY 385 ALQIHILRSHTEGEPFYCNVCGNRFETGKGLVHEHRRREKYPHVQMPRPVPEHLDTYT 444
Db 620 NLKVHLRVHSGEERPFKQOTCNKGFQOLAHLOKHVLYHTGEKPH----- 662
OY 445 SSGLPYGMVPPRKAEEEAATPGGVEKKPLVASTALSATSLTLLSTASGATAPGLP 504
Db 663 ----- 662
OY 505 AFNKFLMKAVERPKNADENTPPGSEGSALSGVAESSTATLMQLSKLMTSLPMSALLTNH 564
Db 663 ----- 662
OY 565 FKSTGSPRLPLCARALGASPSETSKLOOLVEKIDROGAVAVTSAASGAPTTASAPBSSA 624
Db 663 ----- 662
OY 625 SSGPNCVCILRVLSCPRALRHYGQNGGERPKKVCGRAFSTRGNLRAHFHGKASPA 684
Db 663 -----KCOVCHKRRPSTSNLKTHLRLHSGEKPYOCKVCRAKTFQVHLKL---HKRLHT 713
OY 685 ARAQNSCPLCOKFTNAVTLOQAHVHMLGQIIPNGGTALPEGGGAQENSGSTVSGAG 744
Db 714 RRRPHKACQCHKSVYIHCSL---KVHLKGNCRAG-----PAAG----- 748
OY 745 SFPPQOSQDPSPEELSEEBE-----EDEEEEDVTDSDSLAGSGESGGE 791
Db 749 -----LPLEDLIRIMEIERFDISDNADRLMEDSDVDVT-----SMVE 787
OY 792 KAL--SVRGSGSEAS---GAEEVGT--VAAAAATAGKEMD-SNEKTTQOSSLP PPPPPDS 843
Db 788 KETLAVAVRKEKEETSLKVSLOQRMNGNGLLSSGCSLYESSDLSLMLKPLHSPNPLPLVPVKV 847
OY 844 LDQOPRME 851
Db 848 QETVERPMD 855
RESULT 35
E88280
protein eg1-43 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E88280
R:Anonymous, The C. elegans Sequencing Consortium.
S:Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see webistes genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88280
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-543 <STO>
A:Cross-references: GB:chr.II; PIDN:CAA91353.1; PID:93879174; GSPDB:GM00020
A:Note: similar to zinc finger, C2H2 type (4 domains)
C:Genetics:
A:Gene: eg1-43
A:Map position: 2
Query Match 5.0%; Score 264.5; DB 2: Length 543;
Best Local Similarity 25.7%; Pred. No. 9.4e-06;
Matches 108; Conservative 59; Mismatches 191; Indels 63; Gaps 14;
OY 350 NGSGEL--SYGEVWGPLEKPGGRHKRCFAKVGSDALQIHLRSHTEGEPFYCNVCGNR 407
Db 136 NGDEEYIREHGLRPGQTPPGSHKCGVCPKSFSSASGLKQSHINCSLKFPHCLCPKS 195
OY 408 FTTGNLKVHFNHREKY-----PHVQMP--HPVEHLDTYTSS 446
Db 196 YTOFSNLCRRHRRVSDGWTCTPCOSQMPQALTKRHPVCEMTALYKPLMAQLAGLSGAG 255
OY 447 GL-----PYGMSVPRPEKAEDEAATPGGVEKKPLVASTALS--ATESLTLSTAGTA 498
Db 256 GLGSPVRYPHILOMATQRYRPLFLAANPEAYKLMQTTCASPRDABCSGHSSESPPT 315
OY 499 TAB-GLPAFNKFLMKAVERPKNADENTPPGSEGSALSGVAESSTATLMQLSKLMTSLPS 557
Db 316 TEEDVLDLATPKRPTSEMETTSKSDGEDRDSIGSGNDDDDSEAGVLDESSTTTS--- 372
OY 558 WALLTNHFKSTGSPRLPLCARALGASPSETSKLOOLVEKIDROGAV---AVTSAASGAP 613
Db 373 ---TKKRPSTSHITSDILAAPOLGAQALNSTPLGMLORSILNYPAYSPHSLRAMSGAK 428
OY 614 TTSAPAPSSASGPNOCVILRVLSCPRALRL--HYQNGGERPKCKVCGRAFSTRGN 671
Db 429 ASSSPSSSGSKDRYTCKFCQKVF--PRSANLTLRLKRLHTGTGQPYKQCCENSFSISSN 486
OY 672 LRAHF--VGHKASPAARAQNSCPLCOKFTNA-----VTLOQAHVHMLGQIIPNGGTAL 723
Db 487 LQNHVRNIRHNKRNPSLPRHNHNR--QRLNHNSTSTTTTYYHPLRLHL-----PQTSV 538
OY 724 P 724
Db 539 P 539
RESULT 36
S06546
finger protein (clone X1COF.1) - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 01-Dec-2000
C:Accession: S06546
R:Nietfeld, W.; El-Baradi, T.; Mentzel, H.; Pieler, T.; Koester, M.; Poetting, A.; Kno
J. Mol. Biol. 208, 639-659, 1989
A:Title: Second-order repeats in Xenopus laevis finger proteins.
A:Reference number: S05632; MUID:90040698; PMID:2509712
A:Accession: S06546
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-615 <NIE>
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger
Query Match 5.0%; Score 264.5; DB 2: Length 615;
Best Local Similarity 21.4%; Pred. No. 1.1e-05;
Matches 134; Conservative 61; Mismatches 237; Indels 195; Gaps 21;
OY 365 EKPGRHKRCFAKVGSDALQIHLRSHTEGEPFYCNVCGNRFETGKILVHFNHREK 424
Db 140 EKP---FCCSPDCDKCFASSSELNHOQRTHGKRFCSGCKFTNHS---HFAHNO-- 190
OY 425 YPHVQMPHPVPEHLDTYTSSGLPFGMSVPRPEKAEDEAATPGGVEKKPLVASTALSA 484
Db 191 -----MHHG-----EKPFCSKCKGCF 208

```

y 485 TESLLTSTAGIATAPGLPAFNKFLVLMKAVERKNKADENPPGE-GAISGVASSSTA 543
db 209 ASSSDL-----TFRR-----RTHTEKFTSCSECGKCFSNHSHLARR 245
y 544 TLMQ-----LSKIMTSLPSMALLTNHFKSTGSPPLPCARLALGASPESTKLT--QOLV 594
db 246 QMIHNGEKFPCSECGKCFSSSSSLTAHQRTNKKVRFPGCAGCKCFENRSHLTHQMT 305
y 595 EKIDRQGAVAVTSASGAPTTAPAPSSASSGNQVCYICLVLSCPRALRLHYGQHGE 654
db 306 HTGK-----PFSCFCEKRCFSPNSMLARHOMHTGE 337
y 655 RFKCKVNGRAESTRGNLRAHFVGHKASPARAONSCITCKKFTNVAVLQOHVNRHLGS 714
db 338 KPFSCSECGKCFASSSSDLTTHNRHT---TGKPRPSCGCGCYKRSKSLVNHQRTHT-G 392
y 715 QIPNGTALPEGGAOENSGEOSTVGAGSFPPOOOSOPSPHEELSEEEDEEEDEED 774
db 393 EKPFSCSKDKCKASSSELNTHQRTHTGEKAF-----SCSECGKC 432
y 775 VTDEDSLGRSGSGGKKAISVRGDSBPASGAEHEVGTVAATAATGKEMDSNEKTTQSS 834
db 433 FTNRSLSHQMTHTGEKPISC-PECECF-----VSSSLTAHQOQ 473
y 835 LPPPPPSLDQQRPMHQGSSGVLGKKEGK--PERSSSPASALTPBEENAVTLVEEL 892
db 474 AHRMYKPFSC-----LECGKCFSPNSRNFARHOMHTHTG-----KPF 509
y 893 SLQEMARKPEGESSSRK-----ACEVCGQAFPSQALAEHOKTHPKREGPLFTCV 941
db 510 SCSECKKGFSPNSQSLARHQMTHTGECFPCSECGKRFSPNSQSLARHQMTHTGKPF-FSCS 568
y 942 FCRQGFLEBATLKKHMLLAHQVQPPA 968
db 569 ECAGKGFSPNSGLARHQ-MTHTEKPF 594

RESULT 37
\60191
ncogene Evi-1 - human
:Species: Homo sapiens (man)
:Date: 20-Feb-1993 #sequence_revision 17-Apr-1993 #text_change 05-Nov-1999
:Accession: A60191
:Morishita, K.; Parganas, E.; Douglass, E.C.; Ihle, J.N.
ncogene 5, 963-971, 1990
:Title: Unique expression of the human Evi-1 gene in an endometrial carcinoma cell line
:Reference number: A60191; MUID:90326419; PMID:2115646
:Accession: A60191
:Molecule type: mRNA
:Residues: 1-1051 <MOR>
:Cross-references: GB:X54989; NID:g50873; PIDN:CAA8735.1; PID:g50874
:Note: the authors translated the codon CCT for residue 85 as Leu
:Genetics:
:Gene: GDB:EV11
:Cross-references: GDB:119889; OMIM:165215
:Map position: 3q26-3q26
:Keywords: alternative splicing; DNA binding; zinc finger

Query Match 5.08; Score 264.5; DB 2; Length 1051;
Best Local Similarity 18.7%; Pred. No. 1.9e-05;
Matches 171; Conservative 84; Mismatches 236; Indels 425; Gaps 34;

y 368 GGRHKCRFAKVGSGSALQILHRS-HTGERDYKCNVCGNPTTTRGNLKVHPRH----- 421
db 128 GNHVCENCACAVETDPSNLORHNRSHQVGAARAHCPECGKTFATSSGLQKHKIHSSVKP 187
y 422 -----REKYPHV-----OMNPH-PPVEHLDYVIT 444
db 188 FICEVCHKRYTQPSNLCRHKRHMADCRTOIKCKDCGGMFTSTSSLNKHKRPFCEGKNHFAA 247
y 445 SSGLRPGMSVPRPEKAEENAT----- 465

```

Db 248 GGFGGGIGISLPGPAMDKTSVMNNSHANPGLADYFGANRHPGLPTTPAGSFSFSGFL 307

QY 466 PGGVEVEKPKPIVAST---ALSTATESLTLTSTAGT-----ATAGLPAPFNKEFLM---K 513

Db 308 PSLGYRHPPLPIPASSPYKGLSTSEQTNKSQSPIMTHPOLIPATFOIDLKSKHPVGDK 367

QY 514 AVE--PRNKAD-----NTPPGSE-----GSAISVAESSTATLMOL 548

Db 368 PVALQERSSEERPEFEKISDQSESSDLDDVSTPSGSDLETTVSGSDLESIDESDKEFEKEN 427

QY 549 SKL----MTSLPSALLTN-----HFKS 567

Db 428 GKMFKDKVSPLONLASINNKKEYSNHSTFSPLEBOTAVGCAVNDISKAIASTAEYFGS 487

QY 568 TG-----SFLP-----LCARALGASPETSRLQ 592

Db 488 TGLVGLDQKRVKVALPYRSMPLPFPFPAFSQSMYPPRMDLRSLPLKMEQSPGEVAKLQ 547

QY 593 -----LVEKIDRGAAVNTSAAGAPPTTSAPAP-----SSASAG----- 627

Db 548 GSSESEFDLTTRKDEKPLFPVPSPKPYTPATSDQPDLSMGSRSRSGATKLTPEPRKNH 607

QY 628 ----- 627

Db 608 VFGGKGSNVESRPASDGSLOAHNPPTPEMDPIYVEKRRKLTDPLEALKETYLPSBGL 667

QY 628 -----PNO----- 630

Db 668 FHPQFQLPDQRTMMSAIENTNAEKLESFALKPEASELLQSVSMFNFRAPNMLPENLR 727

QY 631 -----CVICLRVLSPPRLRL--HYGQHGGERPKCKVCGRASTRGRLRAHF--VGH 679

Db 728 GKGERYTCRCGCKIF--PRSANLTRHLRTHGEGPYRCQYCCRSFSSISNLQHVNRHIN 785

QY 680 KASPARAQNQSCPIQCKKFTNAVTLQGHVBMHLGQIPNGGALP---EGGAA----- 720

Db 786 KEKPF---KCHLYCFCGQQTNDLRHLKHNENMSGSTATSPSELESTGAILDDKE 840

QY 731 -----QENQSE-----QSTVSGAGSPFOQSQSPPERELSEEEDEDE-- 769

Db 841 DAYFTEIRNFITGNSNHSQSPRNVEERNGS--HFKEKALVPSQNSDLDLDEDEVEDEVLL 899

QY 770 -EEEDV---TDSDLAGSGESGGKKAISVRDSEASG-----AEEVG 811

Db 900 DEDEDEYDITGKGEPEVNTSLHEGNPE-----DDYETSALEMSKTSPVRYKKEEYK 953

QY 812 TVAA-----ATGKENDSNE--KTTQSSLPPEPPPSLDQDQPEQGGSGV----- 857

Db 954 SGLSALDHITHFTDSLKMRKEDNQYSEALSSFTSHVPEELKQPLHKKSKQATAMML 1013

QY 858 -LGKKEGGKPERSS 872

Db 1014 SLSDKESLHSTSHSS 1029

RESULT 38

S03677

C:Species: Mus musculus (house mouse)

C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 01-Dec-2000

C:Accession: S03677

R:Chowdhury, K.; Rohdewald, H.; Gruss, P.

A:Title: Specific and ubiquitous expression of different zn finger protein genes in t

A:Reference number: S03677; MIMD:8905528; PMID:3145103

A:Accession: S03677

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-428 <CHO>

A:Cross-references: GB:X12592; NID:g53132; PIDN:CAA31105.1; PID:g53133

A:Note: The sequence from fig.2 is inconsistent with that shown in fig.1 in having 34

C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

C:Keywords: DNA binding; zinc finger

AL2_HUMAN
SAL2_HUMAN STANDARD; PRT; 1007 AA.
O9Y4G1: O9Y4G1: 15-JUN-2002 (Rel. 41, Created)
15-JUN-2002 (Rel. 41, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Sal-like protein 2 (zinc finger protein SAL2) (Hsal2).
SAL2 OR SAL2 OR KIA0360.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCBI_Taxid=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1).
MEDLINE=97131507; PubMed=8975705;
Kohlhase J., Schu R., Dowe G., Kuehnlein R.P., Jaackle H., Schrodere B., Schulz-Schaeffer W., Kretschmar H.A., Koehler A., Meiler U., Raab-Vetter M., Burkhart E., Engel W., Stick R.; "Isolation, characterization, and organ-specific expression of two novel human zinc finger genes related to the Drosophila gene spalt."; Genomics 38:291-298(1996).
[2]
SEQUENCE OF 11-1007 FROM N.A. (ISOFORM 2).
TISSUE=Brain;
MEDLINE=9734984; PubMed=9205841;
Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.; "Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro."; DNA Res. 4:141-150(1997).
[3]
SEQUENCE OF 541-1005 FROM N.A. (ISOFORM 2).
Morgan J.W., Ford D., Ma Y., Maizel A.L.; "Homo sapiens mRNA for zinc finger protein, SAL2 exon 2."; Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: Probable transcription factor.
-1- SUBCELLULAR LOCATION: Nucleus (Probable).
-1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be produced by alternative splicing.
-1- TISSUE SPECIFICITY: Highest levels in adult brain (in different areas). Lower levels in heart; very low levels in kidney and pancreas.
-1- DEVELOPMENTAL STAGE: In fetal brain exclusively in pontine nuclei.
-1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X98834: CA67351.1: ALT_INIT.
EMBL: AB002358: BAA21638.1: -.
EMBL: AF465630: BAA74188.1: -.
Genew; HGNC:10526; SAL2.
MIM: 602219; -.
InterPro: IPR000822: Znf_C2H2.
Pfam: PF00096: Zf-C2H2; 7.
PROSITE: PS00028: ZINC_FINGER_C2H2_1; 7.
PROSITE: PS50157: ZINC_FINGER_C2H2_2; 7.
Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger; Metal-binding; Repeat; Alternative splicing.
ZNF_FING 373 395 C2H2-TYPE.
ZNF_FING 401 423 C2H2-TYPE.
ZNF_FING 631 653 C2H2-TYPE.
ZNF_FING 659 681 C2H2-TYPE.
ZNF_FING 713 733 C2H2-TYPE.
ZNF_FING 911 933 C2H2-TYPE.
ZNF_FING 940 964 C2H2-TYPE.

FT	DOMAIN	135	139	POLY-GLY.
FT	DOMAIN	158	170	POLY-PRO.
FT	DOMAIN	256	265	POLY-SER.
FT	DOMAIN	764	775	POLY-GLU.
FT	DOMAIN	816	821	POLY-ALA.
FT	DOMAIN	838	843	POLY-PRO.
FT	VARSPLIC	575	581	LPICARA -> FPVLEP (IN ISOFORM 2).
FT	CONFLICT	11	24	OLISDCGFSASEN -> SRLGVPCGEPELG (IN REF. 2).
FT	CONFLICT	547	547	L -> R (IN REF. 2).
FT	CONFLICT	554	554	V -> M (IN REF. 1).
SQ	SEQUENCE	1007 AA;	105135 MM;	92D963225A5459E CRC64;

Query Match 97.9%; Score 5167; DB 1; Length 1007;
Best local similarity 99.4%; Pred. No. 7.3e-219;
Matches 984; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY	16	GEPAELGDASEEDHPQVACCAQFTDPTTEFLAHONACSTDPPYAVITIGGEPNNSSA	75
DB	18	GPSASESGDASEEDHPQVACCAQFTDPTTEFLAHONACSTDPPYAVITIGGEPNNSSA	77
QY	76	SSEPRPEGHNNPQVMDTEHSNPPDSGSSVPTDPFWGPPRRGEESGHFLVAATGTAAGG	135
DB	78	SSEPRPEGHNNPQVMDTEHSNPPDSGSSVPTDPFWGPPRRGEESGHFLVAATGTAAGG	137
QY	136	GGLIASPKIGATPLPEESTPAPPPPPPPPPVSGHLNPLILEELRVLOQRIHQM	195
DB	138	GGLIASPKIGATPLPEESTPAPPPPPPPPPVSGHLNPLILEELRVLOQRIHQM	197
QY	196	QMTEDICQVLLLSIGQTVGAPASBSELPGTASSTKPLPLFSPKVPOTSKTASS	255
DB	198	QMTEDICQVLLLSIGQTVGAPASBSELPGTASSTKPLPLFSPKVPOTSKTASS	257
QY	256	SSSSSSSGAETPKAFFLHYPLHSGHPESAGVGRHKRTAPSPALPGSTDOIASP	315
DB	258	SSSSSSSGAETPKAFFLHYPLHSGHPESAGVGRHKRTAPSPALPGSTDOIASP	317
QY	316	HLAPSTTGILAAOCLGAARGLATASPGLLPKRNGSELGYGEVWGPLEKGGRRKCRP	375
DB	318	HLAPSTTGILAAOCLGAARGLATASPGLLPKRNGSELGYGEVWGPLEKGGRRKCRP	377
QY	376	CAKYFGSALQIHLRSHGERPKYCNVCGNRFTRGNLKAHFNHREKTPHYQNNHPY	435
DB	378	CAKYFGSALQIHLRSHGERPKYCNVCGNRFTRGNLKAHFNHREKTPHYQNNHPY	437
QY	436	PEHLDYVITSSGLPYGMSVPEPEKAEPAATPGGVERKPLVASTALSTESLTLSSTA	495
DB	438	PEHLDYVITSSGLPYGMSVPEPEKAEPAATPGGVERKPLVASTALSTESLTLSSTA	497
QY	496	GTATAPGILPAEFKFLMKAVEPKAKADENTPPGSEGAISGVAESSTATLMQLSKMTSL	555
DB	498	GTATAPGILPAEFKFLMKAVEPKAKADENTPPGSEGAISGVAESSTATLMQLSKMTSL	557
QY	556	PSMALTNHFKSTGSEPLPLCARALGASSETSLOQLVEKIDRGAVAVTAASAPPT	615
DB	558	PSMALTNHFKSTGSEPLPLCARALGASSETSLOQLVEKIDRGAVAVTAASAPPT	617
QY	616	SAPAPSSASSGPNOCVLCRLVLSCPRALRLHYGOHGERPFCKVCVGRAPSTRGNLRAN	675
DB	618	SAPAPSSASSGPNOCVLCRLVLSCPRALRLHYGOHGERPFCKVCVGRAPSTRGNLRAN	677
QY	676	FVGHKASPAARAQNSCPICQKFTNAVTLQOHVBMILGGQIPNGGTALPEGGAAQENGS	735
DB	678	FVGHKASPAARAQNSCPICQKFTNAVTLQOHVBMILGGQIPNGGTALPEGGAAQENGS	737
QY	736	EOSTVSGAGSPPOOQSOQPSPEELSEEEEDDEEDVDDEDSIARGSSGGEKALS	795
DB	738	EOSTVSGAGSPPOOQSOQPSPEELSEEEEDDEEDVDDEDSIARGSSGGEKALS	797
QY	796	VRGDSSEASGAEEVGTAAATAGKEMDSNEKTTQOOSLPPPPPPDLDPQPPQEGSS	855
DB	798	VRGDSSEASGAEEVGTAAATAGKEMDSNEKTTQOOSLPPPPPPDLDPQPPQEGSS	857

QY 856 GVLGKEEGKPERSSSPASALTPEGEATSVTLVEELSLQEMARKKEPGESSSRKACEVCG 915
D 858 GVLGKEEGKPERSSSPASALTPEGEATSVTLVEELSLQEMARKKEPGESSSRKACEVCG 917
QY 916 QAFPSQAALEEHOKTHPKREGPLFTVCFCROGFLEBRATLKKHMLAHNOVPAPRHPGPN 975
D 918 QAFPSQAALEEHOKTHPKREGPLFTVCFCROGFLEBRATLKKHMLAHNOVPAPRHPGPN 977
QY 976 AALSLVPGCSPSTSTGSLSPFRKDDPTIP 1005
D 978 AALSLVPGCSPSTSTGSLSPFRKDDPTIP 1007
RESULT 2
SAL2_MOUSE STANDARD; PRT; 1004 AA.
AC Q9QX96;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
CN Sal-1-like protein 2 (spalt-like protein 2) (Msal-2).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20069074; PubMed=10602995;
RA Kohlhae J., Altman M., Archangelo L., Dikens C., Engel W.;
RT "Genomic cloning, chromosomal mapping, and expression analysis of
Meal-2";
RL Mamm. Genome 11:64-68(2000).
CC -1- FUNCTION: Probable transcription factor.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: Expressed throughout embryonic development. In
CC adult predominantly in brain.
CC -1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ007396; CAB65274.1; -.
DR MGD: MGI:1354373; Sal12.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 7.
DR SMART: SM00355; Znf_C2H2; 7.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
KW Transcription regulation; DNA-binding; Nuclear protein; zinc-finger;
KW Metal-binding; Repeat.
FT ZN_FING 372 394 C2H2-TYPE.
FT ZN_FING 400 422 C2H2-TYPE.
FT ZN_FING 629 651 C2H2-TYPE.
FT ZN_FING 657 679 C2H2-TYPE.
FT ZN_FING 689 711 C2H2-TYPE.
FT ZN_FING 908 930 C2H2-TYPE.
FT ZN_FING 937 961 C2H2-TYPE.
FT DOMAIN 135 139 POLY-GLY.
FT DOMAIN 158 170 POLY-PRO.
FT DOMAIN 256 264 POLY-SER.
FT DOMAIN 762 772 POLY-GLU.
FT DOMAIN 834 840 POLY-PRO.
SQ SEQUENCE 1004 AA; 104975 MW; E46DD2C9C264A9C2 CRC64;

Query Match 87.4%; Score 4613.5; DB 1; Length 1004;

Best Local Similarity 89.0%; Pred. No. 1e-194;
Matches 881; Conservative 25; Mismatches 81; Indels 3; Gaps 3;
QY 16 GEPAEIAGDASEEDHPVYKACCAQFTDPTFEFLAHQNMNSTPPVAVIIGGQENPNSSA 75
D 18 GPSASENGSAEENHPVYKACCAQFSDPTFEFLAHQNSTCTPPVAVIIGGQENPNSSA 77
QY 76 SSEPREGHNPNVMDENSHNPPDSGSVPDPTMGPERGEGSSGHFVAATGTAAGG 135
D 78 SSAPREGHRSQVMDENSHNPPDSGSPDPPTMGPERGEGSSQFLVAATGTAAGG 137
QY 136 GGLILASPKLGATPLPESTPAPPPPPPPPPPPVGGSHLNIPLILEELRVLQROIHOM 195
D 138 GGLILASPKLGATPLPESTPAPPPPPPPPPPPVGGSHLNIPLILEELRVLQROIHOM 197
QY 196 QMEQICROYVLVLLGLOQTVAPASPSLPGTGTAASSTPRLPLFSPDKPVQNSKTASS 255
D 198 QMEQICROYVLVLLGLOQTVAPASPSKLPCTGAASSTPRLPLFSPDKPVQNSKTASS 256
QY 256 SSSSSSSGAETPKQAFPHLYHPLGSQHPFSAGVGRSHKPTPAPAPALPGSTDLIASP 315
D 257 SSSSSSSGAETPKQAFPHLYHPLGSQHPFSAGVGRSHKPTPAPAPALPGSTDLIASP 316
QY 316 HLAFPSTTGLLAOCGLAARGLEATASPGLLKPKNGSGELSTGEVMOPLKPGGRHKCF 375
D 317 HLAFPSTTGLLAOCGLAARGLEAASPGLLKPKNGSGELSTGEVMOPLKPGGRHKCF 376
QY 376 CARVFGSDALQTHLRISITGERPKYCNVGNFTTRGNLKVHHRREKPHYOMNHPV 435
D 377 CARVFGSDALQTHLRISITGERPKYCNVGNFTTRGNLKVHHRREKPHYOMNHPV 436
QY 436 PEHLDVYITSSGFLPYGMSVPEPEKAEAEATPGGVERKPLVASTALSTASTLSTSA 495
D 437 PEHLDVYITSSGFLPYGMSVPEPEKAEAEATPGGVERKPLVASTALSTASTLSTSA 496
QY 496 GTATAPGLPAFNKFLVLMKAVEPKNKADENTPPGSEGAISGVAESSTATLMQSKMTSL 555
D 497 STAVAPGLPFTFNKFLVLMKAVEPKSKADENTPPGSEGAISGVAESSTATLMQSKMTSL 556
QY 556 PSNALLTNHFKSTGSPFLPLCARALGASPSSTKLOQVLEKIDROGAVVATSAAGAPT 615
D 557 PSNALLTNHFKSTGSPFLPLCARALGASPSSTKLOQVLEKIDROGAVVATSAAGAPT 616
QY 616 SAAPSSASSAGPQVCYICLRLVSCPRALRLHYGCHGERPFCKYCGRAFSTRGNLRAH 675
D 617 SAAPSSASSAGPQVCYICLRLVSCPRALRLHYGCHGERPFCKYCGRAFSTRGNLRAH 675
QY 676 FVGHKASPARAONSCPTICQKKFTNAVTLQOHVRLHGGQIPNGCTALPBGCGAAQENG 735
D 676 FVGHKASPARAONSCPTICQKKFTNAVTLQOHVRLHGGQIPNGCTALPBGCGAAQENG 735
QY 736 EOSTVSGAGSPFOQSOQSPPEELSEEEDEDEEDVTDSDSLAGRSESGEKRAIS 795
D 736 EOSTVSGAGSPFOQSOQSPPEELSEEEDEDEEDVTDSDSLAGRSESGEKRAIS 794
QY 796 VRGSEASGAEAEETGYAAATATAGKEMDSNEKTTQOSSLPPPPPSLDQPPQMEGSS 855
D 795 VRGSEASGAEAEETGYAAATATAGKEMDSNEKTTQOSSLPPPPPSLDQPPQMEGSS 854
QY 856 GVLGKEEGKPERSSSPASALTPEGEATSVTLVEELSLQEMARKKEPGESSSRKACEVCG 915
D 855 DVSGAMEEAEALLEGISPMALATQEGESTPLVEELINPEAMKKDGESSGRKACEVCG 914
QY 916 QAFPSQAALEEHOKTHPKREGPLFTVCFCROGFLEBRATLKKHMLAHNOVPAPRHPGPN 975
D 915 QAFPSQAALEEHOKTHPKREGPLFTVCFCROGFLEBRATLKKHMLAHNOVPAPRHPGPN 974
QY 976 AALSLVPGCSPSTSTGSLSPFRKDDPTIP 1005
D 975 AALSLVPGCSPSTSTGSLSPFRKDDPTIP 1004

RESULT 3

SAL3_HUMAN STANDARD; PRT; 1300 AA.
ID SAL3_HUMAN STANDARD; PRT; 1300 AA.
AC Q9BXA9; O9UGH1:
27 15-JUN-2002 (Rel. 41, Created)
27 15-JUN-2002 (Rel. 41, Last sequence update)
27 15-JUN-2002 (Rel. 41, Last annotation update)
28 Sal-1-like protein 3 (zinc finger protein SAL3) (hsSAL3).
28 SAL3.
29 Homo sapiens (Human).
29 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
29 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
29 NCBI_TaxID=9606;
29 [1]
29 SEQUENCE FROM N.A. (ISOFORM 3).
29 Gooding R., Angelicheva D., Blechschmidt K., Snoboda K., Molnar M.,
29 Tourneir I., Kalaydjieva L.,
29 "Exclusion of HSAL3 and refinement of the region for the CCFDN
29 gene";
29 Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
29 [2]
29 SEQUENCE OF 29-1300 FROM N.A., AND ALTERNATIVE SPLICING.
29 MEDLINE=20079154; Pubmed=10610715;
29 Kohlbase J., Hausmann S., Stojmenovic G., Dixkens C., Bink K.,
29 Schulz-Schaeffer W., Altman M., Engel W.,
29 "SAL3, a new member of the human spalt-like gene family, maps to
29 18q23";
29 Genomics 62:216-222(1999).
29 [3]
29 -1- FUNCTION: Probable transcription factor.
29 -1- SUBCELLULAR LOCATION: Nuclear (Probable).
29 -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1, 2, 3 (shown here) and 4; are
29 produced by alternative splicing. Isoforms 1 and 2 lack two zinc
29 finger domains and are the major isoforms.
29 -1- TISSUE SPECIFICITY: Widely expressed in adult with highest levels
29 in heart. Expressed in fetal brain (in neurons of hippocampus,
29 cortex, mediodorsal and ventrolateral thalamic nuclei, putamen,
29 cerebellum and brainstem).
29 -1- DEVELOPMENTAL STAGE: In fetal brain of the 24th gestational week.
29 -1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
29 PROTEINS.
29 -----
29 This SWISS-PROT entry is copyright. It is produced through a collaboration
29 between the Swiss Institute of Bioinformatics and the EMBL outstation-
29 the European Bioinformatics Institute. There are no restrictions on its
29 use by non-profit institutions as long as its content is in no way
29 modified and this statement is not removed. Usage by and for commercial
29 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
29 or send an email to license@sib-sib.ch).
29 -----
29 EMBL; AF347021; AAK18311.1; -;
29 EMBL; AJ007421; CAB65124.1; -;
29 HSSP; P07248; IARE.
29 Genew; HGNC:10527; SAL3.
29 MIM; 605079; -;
29 InterPro: IPR000822; znf_C2H2.
29 Pfam; PF00096; zf_C2H2; 10.
29 SMART; SM00355; znf_C2H2; 10.
29 PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
29 PROSITE; PS00157; ZINC_FINGER_C2H2_2; 8.
29 Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
29 Metal-binding; Repeat; Alternative splicing.
29 ZN_FING 420 442 C2H2-TYPE.
29 ZN_FING 448 470 C2H2-TYPE.
29 ZN_FING 679 701 C2H2-TYPE.
29 ZN_FING 707 729 C2H2-TYPE.
29 ZN_FING 739 761 C2H2-TYPE.
29 ZN_FING 977 999 C2H2-TYPE.
29 ZN_FING 1005 1027 C2H2-TYPE.
29 ZN_FING 1113 1135 C2H2-TYPE.
29 ZN_FING 1141 1163 C2H2-TYPE.
29 DOMAIN 198 201 POLY-ALA.
29 DOMAIN 213 216 POLY-SER.
29 DOMAIN 897 902 POLY-SER.
29 VARSPPLIC 1 133 MISSING (IN ISOFORM 2 AND ISOFORM 4).

FT VARSPPLIC 973 1044 MISSING (IN ISOFORM 1 AND ISOFORM 2).
FT CONFLICT 235 235
FT CONFLICT 593 593
FT CONFLICT 787 787
FT CONFLICT 797 802
FT CONFLICT 808 808
FT CONFLICT 1138 1138
FT CONFLICT 1141 1141
SQ SEQUENCE 1300 AA; 135371 MW; 82CF3BDCB6D59150 CRC64;
Query Match 25.4%; Score 1339.5; DB 1; Length 1300;
Best local Similarity 33.2%; Pred. No. 9.6e-52;
Matches 385; Conservative 115; Mismatches 343; Indels 317; Gaps 43;
QY 16 GEPAEELGGDASEE-----DHPVCAKCCQAFDPTPEFLANQNCSTDPVWVITGGQENPN 71
DB 31 GGEAEADSGPEPSRSGEETVSCERCAEFKMADFLEHQRCSTKLPVLTIV----- 82
QY 72 NSSASPEPRRGHNPNQVMDPENSPPDSGSVPDPTGPERGREGSGHFLVAATCTA 131
DB 83 HEDAPAP-----HEDFPE-----PSPASSPERRASE-----AAEEAG 116
QY 132 AGGGGGIILASPK---LGATPLPEESTPAPPPPPPPPGVSGHLN----- 176
DB 117 AEGAEGARPEVEKAEPEMDAPADTRAPRPPAPAPPTPAGABSTVTLTALLSTKV 176
QY 177 -----IPLILELRYLQCGQHOMQMTEDQCRVLLIG-- 209
DB 177 AYAOFSGARAGSGAGGVAANAAPVLTLEQLMALQDQOIHQLOLEIORSVALMQRP 236
QY 210 -----SLGQTVGAPA-SPSELPGTASTSKPLPLPSPIKPVQTSKTLASSSSSSSS 262
DB 237 PPRPRLSPAAPASAPGAPAPQLPGLAA-----LPL-----SAGAPAAAIAG 277
QY 263 SGAEPTKQAFPHLYHPL-----GSQHPPSAGCVGSHKP-----TPAPSPALFGSTDLT 312
DB 278 SGPAAP--AAFEAGAPLSRPSGASTP---GGPAPPSAPAPASAPAPAPAPAPQSA 332
QY 313 ASPHLAPFSTGLLA-AQCGAARGLEA-----TASPGILKP----- 348
DB 333 ASSQPOSASTPPALAPBSLGAAPGLSPPLPQTSASGVIFPNPLVSIATANALDPLSA 392
QY 349 --KNGSGE--LSYGEVWGLEPKRGKRCRCACAVFSGDSALQIHLRSHTGERPKCNV 403
DB 393 LKMKRKGKPNVAVSEFKASAEDEPFKHKRCACAVFSGDSALQIHLRSHTGERPKCNV 452
QY 404 CGNRFTTRGKLKYNHNRKREKYRHYQANRPHRYPEHLIDVYITSSGLPYGKSVPEKAEEA 463
DB 453 CGNRFTTRGKLKYNHNRKREKYRHYQANRPHRYPEHLIDVYITSSGLPYGKSLPE----- 506
QY 464 ATPGGGVKRPPLVASTALATESLTLTSTAGTATAPGLPAFNFVLMKAAPPKNADE 523
DB 507 -----KPV---TTLWDSKPYLPVPTVSVGLDLPPTVPAHAGVADSPATPAPASRPQ 554
QY 524 NTPPGSE-----GSAISGVAESSATLT----- 545
DB 555 RSPASSECAASLSPGLNHVSGVAGTAESPQSLGCGPVTKAEPVSLPCTNARAGDAPVG 614
QY 546 MQLSKMTSL-----PSMALLTNHFKSTGSPRLDCAALAGASPSSETGLDLYEK 596
DB 615 AASAAPATVDGAPATLSGFGPLPAVVSDFKA--QPPFGGLDMS--QTSSETSLDQLVEN 670
QY 597 IDROGAVAATSAASGAPTTAPAPASSASSSPNOCVILRLVLSCRALRLAHGOGGERP 656
DB 671 IDKK-----MIDPNCVCIVHRVLSGCSALKMHRRTHTGERP 706
QY 657 FCKCKVGAFFSTGRNLRAHVGKASPAARAQNSCPICQKRTTNAVTLQOHVHMLGGQT 716
DB 707 FCKCKICGAFFTTKGNLKHGVRHAKRPPRLRVQHSCLPICQKRTTNAVTLQOHVHMLGGQT 766
QY 717 PARGTALPEGGGAA-----QENGSGQST-----VS 741
DB 767 PN--TPLPEGFODAMDELAYDKNAETLSSYDDMDDENSMEDDAELKDAATPAKPLLS 824

[illegible][illegible]


```

Db 1184 EOLRRFEAAEZOOKRAAAAAAAAAAAQOQTTPQARDQROEGGAGAPPNIMCARPPF 1243
IOY 870 SSSPASALTPGEGETSVTLVEELSLQEMARKEP-----GESSKRAACEVCGQAFPSQA 922
Db 1244 GMEPNLPLEFPATTONCCNMAMQIAQSVPAAPFNLALSGVRGS-TTCGICYKFPCHS 1302
IOY 923 ALEHOKTHREKEGFLTCVCFRCQGFLERAATLKKHMLIAHQVQ 965
Db 1303 ALEIHYRSHTKERP-FKQSCIDRCGFTTGNLKHMLT--HKIR 1342

RESULT 8
SALM_DROVI STANDARD; PRT; 1402 AA.
AC P39806;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeotic protein spalt-major.
GN SALM.
OS Drosophila virilis (Fruit fly).
CC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
CC Muscomorpha; Phylloidea; Drosophilidae; Drosophila.
CX NCBI_Taxid=7244;
XX (1)
XX SEQUENCE FROM N.A.
XX MEDLINE=94139659; PubMed=7905822;
JA Kuehlein R.P., Frommer G., Friedrich M., Gonzalez-Gaitan M.,
JA Weber A., Wagner-Bernholz J.F., Gehring W.J., Jaechle H., Schu R.;
"Spalt encodes an evolutionarily conserved zinc finger protein of
"novel structure which provides homeotic gene function in the head and
"tail region of the Drosophila embryo.";
EMBO J. 13:168-179(1994).
CT -1- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF THE POSTERIOR-MOST
CT HEAD AND THE ANTERIOR-MOST TAIL SEGMENTS OF THE EMBRYO. PROBABLY
CT FUNCTION AS A TRANSCRIPTIONAL REGULATOR. COULD REPRESS THE
CT TRANSCRIPTION OF THE TSH GENE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT BLASTODERM STAGE AND
CC LATER IN RESTRICTED AREAS OF THE EMBRYONIC NERVOUS SYSTEM AS
CC WELL AS IN THE DEVELOPING TRACHEA.
CC -1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
C EMBL: Z27444; CAA81800.1; -.
JR PTR; S42748; S42748.
JR HSSP; P15822; IBOO.
JR FlyBase; FBgn0013137; DvIR\salm.
JR InterPro; IPR000822; znf.C2H2.
JR Pfam; PF00096; zf-C2H2; 7.
JR PRINTS; PR00048; ZINCINGER.
JR SMART; SM00355; znf.C2H2; 7.
JR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
JR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
W Developmental protein; zinc-finger; Metal-binding; DNA-binding;
W Nuclear protein; Transcription regulation; Repeat.
R ZN_FING 432 454 C2H2-TYPE.
T ZN_FING 460 482 C2H2-TYPE.
T ZN_FING 845 867 C2H2-TYPE.
T ZN_FING 873 895 C2H2-TYPE.
T ZN_FING 905 927 C2H2-TYPE.
T ZN_FING 1336 1358 C2H2-TYPE.
T ZN_FING 1364 1386 C2H2-TYPE.
Q SEQUENCE 1402 AA; 153745 MW; 4CDB2DEE384F70AE5 CRC64;

```

Query Match	Similarity	22.1%	Pred. No. 8.5e-26;	Matches 322;	Conservative 134;	Mismatches 383;	Indels 619;	Gaps 52;
QY	2	AHESRRSRLLGVPAGEPEPLGDAEEDH-----POVACKCAQPTDPTFLAHONAC	54					
Db	57	ANASASAGNSPCPSKSPQSHSEDERPERDEQVSEDELYEVSAGOSESLGELE-----NNAD	112					
QY	55	STDPVAVIIGGQENPNSSASSEPRPECHNNDQVMDTEHSNP-----DGSQSV	105					
Db	113	ETN-----ADHNHN-----NNKLYM-----TKPVEHEVQNAUNLNSMP	150					
QY	106	TDPWMPGERBEREESGHFLVATGTAAGCGGILLASPKIGADPLPPESTPAPRPPPP	165					
Db	151	NST-----PPATNAVIA-----RAQFGATPVTLLEAIONMOMATAQF	190					
QY	166	PPPGVSG-----HLNPLILEELVULOORIHOMOMTOIGROVLLG-----	209					
Db	191	AAKIANASGTDNNAAMKQALFLOQTLEFNLDQQQLFQQLIQLOLSQALBNQYKONDE	250					
QY	210	-----SIGOTVGAAPSE-----LPGTASTKPLP	238					
Db	251	ADEELEPEREDGETDYEEERIADMELRQKAEAMASKARQHLINAGVPAANAE-DP	309					
QY	239	LFSP-----IKPVQSKTLASSSSSSSSSG-AEPKQAFILYHPLGSHPEFA--	287					
Db	310	SHQPHRCRLRLKRRKREEDASAKSSGASAKFGEDESSODALNKKIKEMENPLPFQADL	369					
QY	288	-GGVGRSHKPTAPS--PALPGSTDOLIASPHLAPSTGILLAOICIGARGLEATSPG	344					
Db	370	SSITITNHDLPENSIDLLQKIQEVLDS-----ASQ	403					
QY	345	LKPKNSGELSYEVMGP---LEKPGGRHKCFKAYVGSALDILHLSHTGERPYC	401					
Db	404	IL-ANNAADFAFGDKSSDGKGRNEPFKKRQYQKGVGSDALQIHRSHTGERPFC	462					
QY	402	NCGNRFETKGNLVVHFHREKYPHYOMNPRIPEHDLY-----I	443					
Db	463	NVCGSRITTKGNLKVHQRNAQKPPHYVMNATPIREHMDKFPRLLDQMSPDSSPIQSPA	522					
QY	444	TSSGLPYGMSVPEKA-----EEEAATPG-GGVERKPLVASTALSTESTLTL	491					
Db	523	PATGLP-----PPSTSTLTQOMQSPMSFASPAFGLPIGIRPPM-----ELKSL	567					
QY	492	STAGTATA-----PGLPAFNKV-----LMKAVEPKNADE-----NTP	526					
Db	568	GATAGTAGLPIHPFPQMPGIGALKHTHDOSQDMPTDLRKSSGPSRHEEDNIAARLP	627					
QY	527	PGSE-----GSAISGVAESSTATL-----MQ-----LSKLM	552					
Db	628	VKSELMEEEKTEHTMEATRESAMEERLYLEVIKKEKIDEOQMLQIEQMKREPLTATA	687					
QY	553	TSLPSTALL-TNH-EKSTGSEPLPLCAR-ALGASPSSETSK-----LOOL--	593					
Db	688	TPHPQOCLLPTTHAAAKSPRSILPLOCARILSLIMCSHPTTSHNACAVLIGSQTNLIDLP	747					
QY	594	-----VEKIDROGAVVATSAASA-----PTTGA	617					
Db	748	DNPVPTMPQREDEFFAEKFLPNTSKTDHSPTRSPAGHANAHIPRSPFNPIKHEMAAFV	807					
QY	618	PAPSSAS-----SGPMQVCYLCLVLSCPRALRLYHGOH	651					
Db	808	PRHSDNDSMENFTEVSNSETMKLTELMMKKKTSIDPQOCVQCDKVLSCSKSALQNIYRN	867					
QY	652	GGERPKCKAYCGAFTSTRGULRAHFGVGHKASPAARAONSCPIQCKFTINAVTLQOHVRNH	711					
Db	868	TEREPKRCRIGCAFTTKGMLKTHMAVHKIRPVRNPFHQECVCHKKSNALVLOOHIRHL	927					
QY	712	LG-----GQIPGIGTALPBG-----CG	728					
Db	928	TGEPTDLTGEQIOAAEIRDPSPPSMGRGHFNPNFAAAAFHFCAMPEGGAGGPGATGMPGG	987					

QY 729 ----- 728
Db 988 PHNGLGSESSGGLDDNMDCCGGDDFDIDISSEHLNSNDPAATSDRRSSDDFKSLFEQ 1047
QY 729 -----AAOENGSEOSTVSGAGSPFOQOQOP-----SPEE 758
Db 1048 KLRIPTGVVNINSHORPISAAASNPNISGASAPSAPTSSQPKPCSPYRSSCSFPR 1107
QY 759 ELSEEEDEDEEDVDTE--DSLGRSSESGEKAISVKG----- 798
Db 1108 SVS-----ETSGALDLTPRALPPPLASSSSRSRYPOLLVRRRLARSVSSHRCVPMV 1162
QY 799 ----- 798
Db 1163 RALLSSQLPSPVIGIDCLPGLQHHLQOQHLMQOQXVAAAAAQNHHNQOHAALH 1222
QY 799 -----DSEASGAESEVGTVAATAATAGKEMDSNEKTQOSSLPAPPSPDILD 845
Db 1223 QHQEHLRRAQEVQKAAQVQKAAAAAQAQRES-----PQPPPRS-- 1269
QY 846 QPQPMQSSGVLGKEEGKPERSSPSASALTPBG-----EATSVTLVEELS--L 894
Db 1270 -----GESSVGPAPQPNPLISARPPGMPNLPFPATTQNMCMANQIA 1315
QY 895 QEAMRKP-----GESSRKACEVCGQAFPSOALEBHOXTHPKEGPLTVCFCROGF 947
Db 1316 QSVMPAAPNPPLALSGVRGS--TTCGICIKTPPCHSALHHSHTKERP-FKCNICDRGF 1373
QY 948 IERATLKKHMLLAHQVO 965
Db 1374 TTKGNLQKHMLT--HKIR 1389
RESULT 9
2236_HUMAN
ID 2236_HUMAN STANDARD; PRT; 1845 AA.
AC O9UL36; O9UL37;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 236.
GN ZNF236.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=99389731; PubMed=10458916;
RA Holmes D.I., Wahab N.A., Mason R.M.;
RT Cloning and characterization of ZNF236, a glucose-regulated kruppel-
RL like zinc-finger gene mapping to human chromosome 18q22-q23.";
RL Genomics 60:105-109(1999).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A AND B (SHOWN HERE); ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBQUITOUS. EXPRESSION LEVELS ARE HIGHEST IN
CC SKELETAL MUSCLE AND BRAIN, INTERMEDIATE IN HEART, PANCREAS, AND
CC PLACENTA, AND LOWEST IN KIDNEY, LIVER, AND LUNG.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF085244; AAD55329.1; -.

DR EMBL; AF085243; AAD55328.1; -.
DR HSSP; P07248; IARD.
DR Genew; HGNC:13028; ZNF236.
DR MIM; 604760; -.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 30.
DR Prodom; PD000003; Znf_C2H2; 5.
DR SMART; SM00355; ZNF_C2H2; 30.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 30.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 30.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat; Alternative splicing.
FT DOMAIN 37 1801
FT ZN_FING 37 59
FT ZN_FING 66 88
FT ZN_FING 93 115
FT ZN_FING 121 143
FT ZN_FING 153 175
FT ZN_FING 197 219
FT ZN_FING 225 247
FT ZN_FING 253 276
FT ZN_FING 285 308
FT ZN_FING 482 504
FT ZN_FING 510 532
FT ZN_FING 538 560
FT ZN_FING 566 588
FT ZN_FING 657 679
FT ZN_FING 685 707
FT ZN_FING 713 735
FT ZN_FING 741 763
FT ZN_FING 967 989
FT ZN_FING 995 1017
FT ZN_FING 1023 1045
FT ZN_FING 1051 1073
FT ZN_FING 1166 1189
FT ZN_FING 1195 1217
FT ZN_FING 1223 1245
FT ZN_FING 1251 1273
FT ZN_FING 1657 1680
FT ZN_FING 1686 1708
FT ZN_FING 1722 1744
FT ZN_FING 1750 1772
FT ZN_FING 1778 1801
FT VARSPPLIC 1530 1558
FT VARSPPLIC 1559 1845
SQ SEQUENCE 1845 AA; 203659 MW; 2879EA91D0CD3D8 CRC64;
MISSING (IN ISOFORM A).
VGPOEGSGVEALYLENSDKT (IN ISOFORM A).
ELNLTSGSLPSTPTSPSAISRONLYMS -> GSRSVQHS
Query Match 7.1%; Score 376; DB 1; Length 1845;
Best Local Similarity 19.2%; Pred. No. 1,5e-09;
Matches 173; Conservative 99; Mismatches 302; Indels 328; Gaps 29;
QY 364 LEKPGRRHKRCFCAKVFSGDSALQILHLSHTEGRRYKCNVCNRFRTGRNLKVFHRRRE 423
Db 190 IDRGFTYSCPCGKTFQKPSQLTRHRIHTGERPKSCGKAFNOKGALQTHMIKHTG 249
QY 424 KYRNV-----QMNPPVPEHL 439
Db 250 EKPRACAFCPAAFSOKGNLQSHVQRYHSEVKNGPTYNCTECSCEVKSLGSLNTHISKMM 309
QY 440 DYVITSGLPYGMVPRPEKAEEDAAATPGGVERKPLVASTTALSTESLTLLS----- 492
Db 310 -----GGPQNSTSTSTTEAHVLTATLFTGLPLQGTQEAQATSASSQPSQAVSDVIQQL 362
QY 493 -----TSAGTAPRG-----LPAPNKVLMKAVE-----PKKADENT-PP 527
Db 363 ELSEPAVYESQSPQGOQLSTVGINODILQALENGSLSPAPAAHPNDSCAKTSAP 422
QY 528 GSEGSATSGVAESST-----ATLMOLSKLMTSLPS-----WA 559
Db 423 HAQNPDVSVSSNEQTDPTDADEBKDEQSEPKLDKKEKKMKKKSPFLPGSTIRENGVMMH 482

```

OY 560 L-----LTNFK---STGSPFLPLCARALGASPSST-----KLOQ 592
DB 483 VCPYCAKEPRKPSDLVRIIRITHHEKPRFCPOCFRAFVAKSTLAIHTKITHTGAKFCQY 542
OY 593 LVKIDROGAVAV-----TSAS-----610
DB 543 CMKSFSTSGSLKVIIRLHTGVRFAPCPHCKKPRISGHRKTHIASHEKHTLFRKMRHQRK 602
OY 611 ----GAPTTAPAPS-----SSASSGPNOCVIC 634
DB 603 PAKVRKCTKTVNPVDPIDLPQEPILITDLGLIOPIPKNOFOSITNNNFEVNDREDRYKCFYC 662
OY 635 LRVLSCEPRALRLHYGNGGGERPFKCVKCGRASTRGNLRAHFVGHKASPARA-----687
DB 663 HRAVKKRSCHLKOHIRSHTEGKPFKSCGCRGFVSAGVLKAHIFTHTGLKSFKLCINGAF 722
OY 688 -----QNSCPICQKRTMNAVTLQOHVNMHLGGQLPNGSTALPEGGGA 730
DB 723 TTGGSLLRRHNGIHNDLRPYKCPYCKTKTSLNCKKHKMKT-----RYELA 768
OY 731 QENGSDQSTVSGAGSPFOQSQSPSEELSEEEDEEEDVDYDDEDISLAGRSGSGG 790
DB 769 QOLQOHQQAASIDSTVDQSQMOMASTQMOVETESDELPTAELVVAANPEAMLDEPQ---825
OY 791 EKRAISVRGDEEASGAE-----EEVGYAAAAATAGKEMDS-NEKTTQOSSLPPPPP 841
DB 826 ---HVYGTTEEAGIGQQLADLPLEADEDEGFVAPODPLRGHVHVOFEEQSPAQOSFEFAGLP 881
OY 842 DSL-----DQPO--PMEQSSGVVLGKKEGKPERSSSPASALT-----PEG 881
DB 882 QGFVTVDYTHQOQFPFYVQO-----LQDSTLESQALSTSFHQSLQAPSS 928
OY 882 EATSVT--LVEELSLQF-----AMRKEPESSSRKA--CEVCGAFPSQALIEHQKT 930
DB 929 DGMNVTYTRLQOESSQOELDLQAOQSQFLEDNEDQRRSYRCYCNKGFKKSHLKQHVRS 988
OY 931 HKKEGGLFVCPVROGFLERATLKKHMLAHQVQFPAPHGQNTALSLVPCSPSITS 990
DB 988 HNGEKP-YKCKLQGRFVSSGVLKSHE-KTHGTVAFS-----CSVCNMSFTT 1034
OY 991 TG 992
DB 1035 NG 1036

RESULT 10
ID ZN42_HUMAN STANDARD; PRT; 734 AA.
AC P28698; O9UBW2;
DT 01-DEC-1992 (Rel. 24, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 42 (Myeloid zinc finger 1) (MZF-1).
EN ZNF42 OR MZF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
NC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NX NCBI_Taxid-9606;
KN [1]
KP SEQUENCE FROM N.A. (ISOFORM MZFLA).
LA MEDLINE-91317761; PubMed-1860835;
LA Hromas R., Collins S.J., Hickstein D., Raskind W., Deaven L.L.,
LA O'Hara P., Hagen F.S., Kaushansky K.;
LA "A retinoic acid-responsive human zinc finger gene, MZF-1,
LA preferentially expressed in myeloid cells.";
LA J. Biol. Chem. 266:14183-14187(1991).
LN [2]
NP SEQUENCE FROM N.A. (ISOFORM MZFLB/C).
NC TISSUE-Bone marrow;
CX MEDLINE-20432092; PubMed-10974541;
LA Peterson M.J., Morris J.F.;
LA "Human myeloid zinc finger gene MZF produces multiple transcripts and
LA encodes a SCAN box protein.";
```

```

RL Gene 254:105-118(2000).
CC -1- FUNCTION: MAY BE ONE REGULATOR OF TRANSCRIPTIONAL EVENTS DURING
CC HEMOPOIETIC DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; MZF1A AND MZF1B/C (SHOWN HERE);
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN DIFFERENTIATING
CC MYELOID CELLS.
CC -1- INDUCTION: BY RETINOIC ACID.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 SCAN BOX.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
DR EMBL; M58297; AA55898.1; -
DR EMBL; AF055078; AAD55810.1; -
DR EMBL; AF055077; AAD55809.1; -
DR EMBL; AF161886; AAF80465.1; -
DR PIR; A40751; A40751.
DR HSSP; P08047; 1SP2.
DR TRANSFAC; T00529; -
DR Genew; HGNC:13108; ZNF42.
DR MIM; 194550; -
DR InterPro; IPR003309; Treg-SCAN.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 13.
DR Pfam; PR02023; SCAN; 1.
DR PRINTS; PR00048; ZNCFINGER.
DR PRODOM; PD000003; Znf_C2H2; 7.
DR SMART; SM00431; LER; 1.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS50804; SCAN_BOX; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat; Alternative splicing.
FT DOMAIN 44 125 SCAN BOX.
FT DOMAIN 310 321 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 356 731 ZINC FINGERS.
FT ZN_FING 356 378 C2H2-TYPE.
FT ZN_FING 384 406 C2H2-TYPE.
FT ZN_FING 412 434 C2H2-TYPE.
FT ZN_FING 440 462 C2H2-TYPE.
FT ZN_FING 463 484 GLY/PRO-RICH.
FT ZN_FING 483 507 C2H2-TYPE.
FT ZN_FING 513 535 C2H2-TYPE.
FT ZN_FING 541 563 C2H2-TYPE.
FT ZN_FING 569 591 C2H2-TYPE.
FT ZN_FING 597 619 C2H2-TYPE.
FT ZN_FING 625 647 C2H2-TYPE.
FT ZN_FING 653 675 C2H2-TYPE.
FT ZN_FING 681 703 C2H2-TYPE.
FT ZN_FING 709 731 C2H2-TYPE.
FT VARSPLIC 1 249 MISSING (IN ISOFORM MZFLB/C).
FT VARSPLIC 250 257 EAGGIFSP -> MNGPLVYA (IN ISOFORM
FT MZFLB/C).
FT CONFLICT 304 305 AL -> RV (IN REF. 1).
SQ SEQUENCE 734 AA; 82036 MW; 2BE7D69B18F29437 CRC64;

Query Match 6.1%; Score 321; DB 1; Length 734;
Best Local Similarity 20.7%; Pred. No. 1.5e-07;
Matches 196; Conservative 74; Mismatches 309; Indels 368; Gaps 37;

OY 80 RPEGHNNPOVMDT-----EHSNPPDSGSSVPTPTWPERGERGESSGHFLVAATCTANG 133
DB 74 RPEVSKRQMLLELVLEQFLGALPPEIQARVQ-----GQRPGSPEEAALVDGLRRPBG 127
```

134 GG-----GGLLASPKLGAT---PLPEESTPAPPPPPPPPPGVSGHNLPLILE 182
128 GPRRWVTVVOVGQEVLESEKMEPSSFOPLPETEPTPEPTPPGPTPP----- 171
183 ELRVLOQKQI-----HOMQTEQICQVILLGSGQVGVGAPASPELPGTGAASSTPL 237
172 --RTWQESPLGLQVKEEVEYED-----SDFLESGPLATQESV 208
238 PLFSPKIPVQTSKTLAASSSSSSSSGAEPTPKQAFPHLYHPLGSHQPSAGVGRSHKPT 297
209 PTLPEEAORCGTIVL--DQIFPHSKTGPEGPSWR-----EHRALMHE--EAGCIF----- 255
298 PAPSPAL-----PGSDQLASPHLAPSTTGL--LAQCGAANGLEMTASPLK 347
256 -SPGALDGLSGISAGGS-----VSPHLHVPMDIGMAGLSGQISPSR--EGGFALLL 307
348 PKNGSGELSYGEV---MGPL-----EKPGGRH-----KCRFCARVFGS 382
308 PSDLASEDPTDEDPGRGVPRALITTRMSRGRHSRGRPRSTGGVYRGRCOVCGKVFISQ 367
383 DSAIQILHLSHSGERYKCNVGCNRFRTTGNLKVHFRHREKYPHYQMPHPVPHLDVY 442
368 RSNLIRHQKIHTEGERPFVSCSGRSFSSHLRHLQTHTEERP-----FV 413
443 ITSSGLPYGMSVPEKAEEAATPRGGVERKPLVASTALSTESLITLSTAGTAPG 502
414 CGDGGGFG---VRSARLEHRVHTG---EOPFRCAEGCGQSRORNSLLQHORIHDDPPG 467
503 LPAFNKFVLMKAVERPKKADENTPRGSESAISGVAESSTATIMOLSKIMTSLPWSALLT 562
468 -----FGAKPRAPGAP----- 480
563 NHFKSTGSPPLPLCARALGASPSETSKLQQLVEKIDROGAVNVTSAASCAPTTSAPAPSS 622
481 ---PPGPFPSCSECR-----ESFARRAVLLEHQVH-----TG 509
623 SASSGGNOCVICLRVLSCPRALRYHGOHGGERPKCKVCVGRAFSTRGNLRAHFVGHKAS 682
510 DKSFG---CVEGGERGRRSVLLOHRRVHSGERRPACACGGSFRRRSNLTQHRITHTGE 566
683 -PAAQAQNSCPICQKKTNAVTLQOHVHMLGGQIPNGSTALPEGGAQENSGEOSTVS 741
567 RPPA-----CAEGKAPRQRTPLQHLRVHTGEK---PEACEGCG----- 603
742 GAGSFPQOQSQSPSEELSEEEDEEEDVDDESLAGSGSGKATSVRQDSE 801
604 -----QRESQRLKLRHQRTHT-----GKR----- 623
802 EASGAEVEGTVAAATATAGKEMDSNEKTQSSLPPEPPPSLDQPPMEQSGSVLGK 861
624 -----PYCGEGCL----- 632
862 EEGKBERSSSPASALTPEGEATSVLVEELSLQEAARKREPGSSSRKACEVCGQAPPSQ 921
633 -----GFTQVSRLT--EHQRIHTGRRPF--ACPECGQSFRQH 665
922 AALEHQKTPKKEGRPLTCVCRQGFLEKATLKKHMLLAHQOPRA 968
666 ANLTQHRITHTGERP-YACRPGCKAFRQPTLQH-LRTHRRKPPA 710

RESULT 11
Y296_HUMAN
ID 1296_HUMAN STANDARD; PRT: 1829 AA.
AC 015015;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein KIAA0296.
GN KIAA0296.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC -I- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -I- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -I- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL: AB002294; BAA20756.1; -;
DR InterPro: IPR000822; Znf.C2H2.
DR Pfam: PF00096; Zf-C2H2; 31.
DR PRINTS: PR00048; ZNCFINGER.
DR SMART: SM00355; Znf.C2H2; 30.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 30.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 29.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT DOMAIN 8 1783
FT ZN_FING 8 31
FT ZN_FING 48 70
FT ZN_FING 75 97
FT ZN_FING 239 261
FT ZN_FING 266 288
FT ZN_FING 294 316
FT ZN_FING 374 396
FT ZN_FING 401 424
FT ZN_FING 465 487
FT ZN_FING 492 514
FT ZN_FING 575 597
FT ZN_FING 617 639
FT ZN_FING 644 666
FT ZN_FING 821 843
FT ZN_FING 848 870
FT ZN_FING 881 904
FT ZN_FING 958 980
FT ZN_FING 1052 1074
FT ZN_FING 1079 1101
FT ZN_FING 1203 1225
FT ZN_FING 1230 1252
FT ZN_FING 1258 1280
FT ZN_FING 1299 1321
FT ZN_FING 1326 1348
FT ZN_FING 1364 1386
FT ZN_FING 1557 1579
FT ZN_FING 1585 1607
FT ZN_FING 1677 1699
FT ZN_FING 1704 1726
FT ZN_FING 1732 1754
FT ZN_FING 1761 1783
SQ SEQUENCE 1829 AA; 200760 MW; B20C482B38684895 CRC64;
Query Match 6.0%; Score 318; DB 1; Length 1829;
Best Local Similarity 20.2%; Pred. No. 5e-07;
Matches 230; Conservative 108; Mismatches 346; Indels 452; Gaps 52;
QY 28 EDHPD--VCARKCAQPTDTEFLAHQNACTDPPVYVIGGQENPNNSASSEPRP----- 81

```

2  EDTPRLSCSDCGRHPPSLPELSRHRELHPSP-----NQDSEADSLPRRYRCQ 51
82  ---BEHNNPQVNDTEHNSPPDSSSVPTDTPWGERGEESGHFLVATGTAAGGGGI 138
52  QCGRYRHP-----GSLV-----NHRRTHTGTPCTTCG-----81
139  ILASPKLATPLRPES---TPAPP-----PPPPPPPVGGSGHLNTPLELKLVLQ 189
82  ---KQSPNPAKLSHMTTHAPEGRRHRRPRRKEATP-----HLQ-----118
190  ROIHOMQTEQICROVLLLSGLQGTWGAAPSSELPGTASSSTPKPLPLFSPKPVQTS 249
119  -----GETVSTDSWQQRIGS-----SEGWNQTK-----HTE 145
250  KTLASS-----SSSSSSSGAETPKQAFPHLHPLGSOHP-----FSAGVGRSHK-----PTP 298
146  ETPDCEVYDPDPAASGTWEDLPTROREGL-----ASHRPEDGAGWGPSTNSARAPLP 200
299  APSPALPSTQDLIASPHLAFPTTGILAAQCLGAAGLEATASFGLLPKNGSGELSYG 358
201  IPASLSLNLLEYLAESVNF-----TG-----GQETQSP-----231
359  EVMGPLEPFGGHNKRCFKAKVNGSDSALOIHLSHTEBRPYKCNVCNRFTRGNLKVHF 418
232  ---PAEBE-RRYKCSQCKTKYKHAOSLTGNHROSHT-LGIYPCALCFKFEFSMIMALKNHS 285
419  HRHREKYHVQNNPHVPHLDVITSSGLPYGKSVPEKAEAEATPGGVERKPLVAS 478
286  RL-----HAQYRPHCHP-----CPRVPRRLRELLEHQQSHGEOEPR-----324
479  TTALSTATESLTLTSSAGTATAPGLPAFNKFLVAKAVEPKKAKADENTPPGSGAISGVA 538
325  -----WEEKMPTTNGHT-----DESSODOLPSAQ--MLNGSA 355
539  ESSATLMLQSLMTSLP-----SMALLTNHKS--TGSEPLPLCARALCASSS 585
356  ELSTSGELEDSELETRPRPCDCCGRTYRHAGSLINHRKSHQGTGYPPSLCKOLFNMAA 415
586  ETSKQLOLVEKIDROGAVAVTASAGAPTTSAAPASASSSGPNQVYICLRVLSCPRALR 645
416  LKNHVR--AHNHRPGV-----GENGQPSV-PPAP-----LL 444
646  LHYGONGE-----RPFKVCVGRASTRGNLPAHFVGHKASPARAONSCPTCK 696
445  LAETTHKHEEDPTTLDNHRPYKCEGGRAYRHRKSLVNHRSHT-----GEYQCSLCP 499
697  KFTNAVTLQOHVHMLGCGIIPNGTALPEGGAQENGSEOSTVSGAGSPQOQSOQ--P 754
500  KYPNIMALRNHVRHC-----KAARRSADIGAGARSHLKVLELPP 539
755  SPEERLESEEBED-----EEEBEDVT-----DEDSLGRGSESG- 789
540  DPVEAEAPRHTQDHYVCKHEEEDITTPADKTAANHICSLGLFEDAESLERHGLTHGA 599
790  GEK-----ALSVR-----GD-----799
600  GEKENSRTETTSPPRAPEFCROCKSKYRHSGLINHRQTHQTFSCGACAKHFTMAAM 659
800  -----SEASGAEVEVGTVAATAAGKEMDNSEKTTQOSLPLPPPP 841
660  KNHLRHRHSRRSRHRKRKRGASGGRG--AKLLAESWTRLEENE-----GLESPODP 711
842  -----DSLDPQRPWEGSGSVLGGKEEGKRP-----RSSSPASA 876
712  SGSEPHGAEGNLESDDCLQAESEGGKCGLEBDETHQGDKESSGTGGLERKDALSLDN 771
877  LTPRGE-----ATSVTLVEELSLQEAAMKPEGESS-----SKKACEVCGQA 917
772  LDIPGEEGGGTHFCDSLTVGVD--DQKPRATGPNSSSHSANAVTGWAQAHNTCSDCGHS 829
918  FPSOAALEHOKTHPRKEGFLFTCVCFROGFLERATLKKHMLLAHQ-----VOPF 967

```

```

DB 830  FPHATGILLSHRPCHP--GIYQCSLCPKFPDSLPAIRSH--FQNNRPGEATSAQPF 881
RESULT 12
ID 2341_HUMAN STANDARD; PRT; 773 AA.
AC O95YN7; O96ST5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein 341.
GN ZNF341.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babage A.K., Baggaley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levasalho M.H., Leverkus M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McMurtry A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prachalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilmink L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [2]
RP SEQUENCE OF 227-773 FROM N.A.
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Watanabe M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuo Y.,
RA Ninomiya K., Iwayanagi T.;
RT "MDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -I- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -I- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL050349; CAC14087.2; -
CC EMBL; AK027550; BAB55193.1; -
CC HSSP; P07248; ZADR.

```


KW	Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KM	Nuclear protein; Repeat.
FT	DOMAIN 57 520 ZINC FINGERS.
FT	ZN_FING 57 79 C2H2-TYPE.
FT	ZN_FING 85 107 C2H2-TYPE.
FT	ZN_FING 163 186 C2H2-TYPE.
FT	ZN_FING 274 296 C2H2-TYPE.
FT	ZN_FING 302 324 C2H2-TYPE.
FT	ZN_FING 498 520 C2H2-TYPE.
FT	CONFLICT 232 MISSING (IN REF. 2).
FT	CONFLICT 436 O -> E (IN REF. 2).
ISO	SEQUENCE 722 AA; 76876 MW; B19DA7B148BC45B CRC64;
Query Match	5.7%; Score 303; DB 1; Length 722;
Best Local Similarity	22.4%; Pred. No. 9e-07;
Matches 164; Conservative	70; Mismatches 271; Indels 228; Gaps 27;
2Y	348 PRNGSGELSYGEVMGLERKGGRI-KCRFCAKVFGSDSALOHLHSHTEPERKCVNCGN 406
Db	33 PAVSAGSLMGAAVSWSESRAGERFPCVCGKRERFNSTLALHRAHPGAQAFQCPHCGH 92
Db	407 RFTTGNLKVHFHREKYPHYQMNPHVPEHLDYVITSSGLPYGMSVPEKAEAEATP 466
Db	93 RAAQALRLSHLTHQPERP----- 112
2Y	467 GGGVERKPLVASTTALSATESLTLSTAGTATAPGLPAFNKFLMKAVEPKKADENTP 526
Db	113 -----RSP--AARLLELEERALLREARLGRARRSG-----GMOAT 146
2Y	527 PGSESSAISGVNESSATATLMQSLKMTSLPSWALLTNHFKSTGSPPLPCARALGASPSF 566
Db	147 PATEGLARPOADSS-----AFRCPCYCKKERTS-AE 177
2Y	587 TSKLQOLVEKIDROGAVAVTSA-----ASGAP--TTSAPAPSSASSGPN 629
Db	178 RERHLHLHRPKKCGICSSSOEELHLSLTAHGAPERPLATSAAPRPQDPQPPFQ 237
2Y	630 QCVICLRVLSCPRALRLHYQGNGER-----PFCKVCGRAFSTRGN 671
Db	238 -----PEPRSVQDPREPEREREATPTPAPAPREPPAPREFRCQVCGSQFTQSWF 288
2Y	672 LRAHVGHKASPAARQNSQIKFTNAVTLQOHVNM-----LGQITNGGAL-- 723
Db	289 LKGNHRKHAS---FDHACPVCGRCRKEPWFLLNMKVHNSKIGPLRAAPGAPSGPARP 344
2Y	724 --PEGGAQOENGSEOSTVSGAGSFPOQSQSPSEBELSEEEDEDEEDVTEDSL 781
Db	345 QPRLGLALAYEPLRALLLAPPT-PAERREPPSLGLYSLRAGGRP-----NKGAE 397
2Y	782 AGRGSESG-----EKAISVAGDSEASGAEENVTAANA---TAGKEMDSNEKTTQSS 834
Db	398 PGPGRSFGFRPLSSALPARARRHRAEPEEEVEVLEAOEFTWARGRSLGS-----LAS 451
2Y	835 LPPRRPDLDPQRPWEGSSGVLGKKEGGKPERSSSPASALPREGESTVIVELS- 893
Db	452 LHRP-----GEGPHSASAAGA---OARSTATQENGL 482
2Y	894 LOEAMRKPERGESSSRKACVCGAPRSQALIEHQKTHPKRGPLTCVFCROGFLEATL 953
Db	483 LVGGRTRREGRGATGKDCFCGCSKPSAHNLKVLRYHTGERP-YKCHNCDVAGQGSSL 541
2Y	954 KKHMLAHQOV-----PRAP-----HGFONTAALSLVP-----GCSPTSTSG 992
Db	542 KYH-LQRIHREORSAGPGRPPRPPOSGASAPQAKPSOPATWEGASSPPRSSG 600
2Y	993 LSPFRKDDPTIP 1005
Db	601 AGPGSR-KPASP 612
RESULT 14	
N84_HUMAN	STANDARD; PRT; 738 AA.
D	ZN84_HUMAN

AC	P51523; Q9NNX7; Q9UC17; Q9UC18;
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Zinc finger protein 84 (Zinc finger protein HPR2).
GN	ZNFR84.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Placenta;
RA	MEDLINE=89377476; PubMed=2505992;
RA	Bellefroid E.J., Lecocq P.J., Benhida A., Poncelet D.A.,
RT	Belavay A., Martial J.A.;
RT	"The human genome contains hundreds of genes coding for finger
RL	proteins of the Kruppel type.";
RL	DNA 8:377-387(1989).
RN	[2]
RP	SEQUENCE FROM N.A.
RA	Grimaldi G.;
RL	Submitted (Feb-2002) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE OF 1-257 FROM N.A.
RC	TISSUE=Teratocarcinoma;
RA	MEDLINE=92051312; PubMed=1945843;
RA	Rosati M., Marino M., Frenze A., Tramontano A., Grimaldi G.;
RT	"Members of the zinc finger protein gene family sharing a conserved N-
RT	terminal module.";
RL	Nucleic Acids Res. 19:5661-5667(1991).
CC	-1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC	-1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC	-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC	FINGER PROTEINS.
CC	-1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	at the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M27878; AAA79359.1; -
DR	EMBL; X60156; CAB94232.2; -
DR	HSSP; P08047; ISP2.
DR	Genew; HGNC:13159; ZNF84.
DR	InterPro; IPR001909; KRAB.
DR	InterPro; IPR000822; Znf_C2H2.
DR	Pfam; PF00096; zf-C2H2; 19.
DR	Pfam; PF01352; KRAB; 1.
DR	PRINTS; PR00048; ZINC-FINGER.
DR	Prodom; PD000003; Znf_C2H2; 14.
DR	SMART; SM00349; KRAB; 1.
DR	SMART; SM00355; Znf_C2H2; 19.
DR	PROSITE; PS50805; KRAB; 1.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 19.
KW	Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW	Nuclear protein; Repeat.
FT	DOMAIN 8 79 KRAB.
FT	DOMAIN 207 733 ZINC FINGERS.
FT	ZN_FING 207 229 C2H2-TYPE.
FT	ZN_FING 235 257 C2H2-TYPE.
FT	ZN_FING 263 285 C2H2-TYPE.
FT	ZN_FING 291 313 C2H2-TYPE.
FT	ZN_FING 319 341 C2H2-TYPE.
FT	ZN_FING 347 369 C2H2-TYPE.
FT	ZN_FING 375 397 C2H2-TYPE.
FT	ZN_FING 403 425 C2H2-TYPE.
FT	ZN_FING 431 453 C2H2-TYPE.

FT	ZN_FING	459	481	C2H2-TYPE.
FT	ZN_FING	487	509	C2H2-TYPE.
FT	ZN_FING	515	537	C2H2-TYPE.
FT	ZN_FING	543	565	C2H2-TYPE.
FT	ZN_FING	571	593	C2H2-TYPE.
FT	ZN_FING	599	621	C2H2-TYPE.
FT	ZN_FING	627	649	C2H2-TYPE.
FT	ZN_FING	655	677	C2H2-TYPE.
FT	ZN_FING	683	705	C2H2-TYPE.
FT	ZN_FING	711	733	C2H2-TYPE.
FT	CONFLICT	62	E -> Q (IN REF. 2 AND 3).	
SO	SEQUENCE	738 AA;	85458 MW;	55268AID98179EED CRC64;
Query Match				
Best Local Similarity		5.7%;	Score 301.5;	DB 1; Length 738;
Matches 133; Conservative		21.5%;	Pred. No. 1.1e-06;	
		67;	Mismatches 235;	Indels 185; Gaps 18;
OY	365	EKPGRHNCRCFCAKVFSGSDALQIHLRSHTGPRYKCNVCNRRFTTGNLKVHFRHR-E	423	
Db	260	EKP---YNSOGCKAFSSQKLSQTSLSHQRTHTGKPYECGECGCKAFSRKSHLISHWRTHTGE	316	
OY	424	KYPHVQMPHPVPHLDIVYITSSGLPYGMSVPRPEKAEDEAATPGGCVERRKPLVASTLAUS	483	
Db	317	K-----PYGCNCGRAFSER-----	331	
OY	484	ATESLTLSTAGTATAPGLPAFNKFLVLMKAVEBPKNKADENTPPGSEGSALSGVAESSTA	543	
Db	332	-----SNLTINHQRITHTGKPEECRECKAFSRKS-----	360	
OY	544	TLMOQLSKMTSLPSMALLTNHFKSTGSPPLD-CARAL-----GASPSSET	587	
Db	361	-----OLVTHNHTHTGTRKPEGSCDCKRAFFKSELSLRHOTHTGKPYEC	405	
OY	588	SKLOOLVEKIDROGAAVAVSASGAPRTSAPAPSSASSGPRQCVICLAVLSCPRALRLH	647	
Db	406	SECRKAFR--ERSSLINHQRITHTG-----EKPRGCIQCGKAFSGKSHLISH	449	
OY	648	YGOHGERPFCCKVCGRAFSTRGNLRAHFVGHKASPARAQNCSPTCKKFTNAVTLQOH	707	
Db	450	QMTHTGKFPFICKCGKAFSRKSOL---YRHQRTHTGKPYECSECGKAFSEKSLITNH	505	
OY	708	VAMHLGGQIPNGSTALPREGGAOENSGSEOSTVSGAGSPFOOQSOQSPPEELSEEEEEE	767	
Db	506	QRH-----TGEKPYVSCGKAFCKSHLISHQRTHTGKPYE	544	
OY	768	DEEEDVDVDEDSLAGRSGSGEKAISVRGDSERASGAEEVGVYAAATAGKENDSNE	827	
Db	545	CSECGKAFGEKSLATQRTHTGKPYECR-DCEKAFSOKSOLMT-----HQRTHTGE	596	
OY	828	KTTQGSLLPPRPDSDOPQPMGSGSGVLGKREGGKPERSSSPASALTPREGATSVT	887	
Db	597	K-----PEYCSLCRKAFFKSELSLRHTHTGKPYECNECKRAFR-----	637	
OY	888	LVEELSLQEAARKPERGESSSRKACEVCGQAFPSQALAEHQHTHPKRGPLETCVFCROGF	947	
Db	638	--EKSSLINHQRITHTGKPRE--CSECGKAFSRKSHLIRHQRTHTGKPR-YGSECKRAFF	692	
OY	948	LERATLKKHMLLAHQVQPF	967	
Db	693	SQKSQSLVNHQRT-HTGKPY	711	
RESULT 15				
ID	XFIN_XENLA	STANDARD;	PRT;	1350 AA.
AC	P08045;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	zinc finger protein xfin.			
GN	XFIN.			
OS	Xenopus laevis (African Clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC	Xenopodinae; Xenopus.
OX	NCBI_TaxID=8355;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Embryo;
RX	MEDLINE=88082679; PubMed=2826129;
RA	Ruiz I Altaba A.; Perry O'Keefe H.; Melton D.A.;
RT	"Xfin: an embryonic gene encoding a multifingered protein in
RT	Xenopus.";
RL	EMBO J. 6:3065-3070(1987).
RN	[2]
RP	CHARACTERIZATION.
RX	MEDLINE=94021366; PubMed=7692399;
RA	Andreazoli M.; de Lucchini S.; Costa M.; Baracchi G.;
RT	"RNA binding properties and evolutionary conservation of the Xenopus
RT	multifinger protein Xfin.";
RL	Nucleic Acids Res. 21:4218-4225(1993).
RN	[3]
RP	STRUCTURE BY NMR OF FINGER 31.
RX	MEDLINE=89346749; PubMed=2503871;
RA	Lee M.S.; Gipeert G.P.; Soman K.V.; Case D.A.; Wright P.E.;
RT	"Three-dimensional solution structure of a single zinc finger DNA-
RT	binding domain.";
RL	Science 245:635-637(1989).
RN	[4]
RP	STRUCTURE BY NMR OF A FINGER.
RX	MEDLINE=89378224; PubMed=2506074;
RA	Lee M.S.; Cavanaugh J.; Wright P.E.;
RT	"Complete assignment of the 1H NMR spectrum of a synthetic zinc
RT	finger from Xfin. Sequential resonance assignments and secondary
RT	structure.";
RL	FEBS Lett. 254:159-164(1989).
CC	-I- FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSLATIONAL
CC	REGULATION PROCESSES.
CC	-I- SUBCELLULAR LOCATION: Cytoplasmic.
CC	-I- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH
CC	AS NEURAL RETINA CONES.
CC	-I- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND
CC	EMBRYOGENESIS.
CC	-I- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.
CC	-I- PTM: PHOSPHORYLATED.
CC	-I- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC	FINGER PROTEINS.
CC	-I- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X06021; CAA29425.1; -
DR	PIR; S00647; S00647.
DR	PDB; 1ZNF; 15-OCT-91.
DR	InterPro; IPR001909; KRAB.
DR	InterPro; IPR000822; znf_C2H2.
DR	Pfam; PF00096; zf-C2H2; 36.
DR	Pfam; PF01352; KRAB; 1.
DR	PRINTS; PR00048; ZINCFINGER.
DR	ProDom; PD000003; znf_C2H2; 20.
DR	SMART; SM00349; KRAB; 1.
DR	SMART; SM00355; znf_C2H2; 35.
DR	PROSITE; PS50805; KRAB; 1.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 35.
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 37.
KW	Zinc-finger; Metal-binding; RNA-binding; Repeat; 3D-structure;
KW	Phosphorylation.
FT	DOMAIN 1 58 KRAB.
FT	DOMAIN 108 298 ZINC-FINGERS I.
FT	DOMAIN 326 488 ZINC-FINGERS II.

DOMAIN	503	721	ZINC-FINGERS III.
DOMAIN	750	940	ZINC-FINGERS IV.
DOMAIN	988	1066	ZINC-FINGERS V.
DOMAIN	1136	1298	ZINC-FINGERS VI.
ZN_FING	108	130	C2H2-TYPE.
ZN_FING	136	158	C2H2-TYPE.
ZN_FING	164	186	C2H2-TYPE.
ZN_FING	192	214	C2H2-TYPE.
ZN_FING	220	242	C2H2-TYPE.
ZN_FING	248	270	C2H2-TYPE.
ZN_FING	276	298	C2H2-TYPE.
ZN_FING	326	348	C2H2-TYPE.
ZN_FING	354	376	C2H2-TYPE.
ZN_FING	382	404	C2H2-TYPE.
ZN_FING	410	432	C2H2-TYPE.
ZN_FING	438	460	C2H2-TYPE.
ZN_FING	466	488	C2H2-TYPE.
ZN_FING	503	525	C2H2-TYPE.
ZN_FING	531	553	C2H2-TYPE.
ZN_FING	559	581	C2H2-TYPE.
ZN_FING	587	609	C2H2-TYPE.
ZN_FING	615	637	C2H2-TYPE.
ZN_FING	643	665	C2H2-TYPE.
ZN_FING	671	693	C2H2-TYPE.
ZN_FING	699	721	C2H2-TYPE.
ZN_FING	730	772	C2H2-TYPE.
ZN_FING	778	800	C2H2-TYPE.
ZN_FING	806	828	C2H2-TYPE.
ZN_FING	834	856	C2H2-TYPE.
ZN_FING	862	884	C2H2-TYPE.
ZN_FING	890	912	C2H2-TYPE.
ZN_FING	918	940	C2H2-TYPE.
ZN_FING	988	1010	C2H2-TYPE.
ZN_FING	1016	1038	C2H2-TYPE.
ZN_FING	1044	1066	C2H2-TYPE.
ZN_FING	1136	1158	C2H2-TYPE.
ZN_FING	1164	1186	C2H2-TYPE.
ZN_FING	1192	1214	C2H2-TYPE.
ZN_FING	1220	1242	C2H2-TYPE.
ZN_FING	1248	1270	C2H2-TYPE.
ZN_FING	1276	1298	C2H2-TYPE.
STRAND	1045	1045	
STRAND	1052	1052	
HELIX	1056	1065	
TURN	1066	1066	
SEQUENCE	1350	AA: 155804	MM: 27F10AB0851E0AD8 CRC64;

```

Query Match: 5.68; Score 296.5; DB 1; Length 1350;
Best Local Similarity 21.38; Pred. No. 3.2e-06;
Matches 131; Conservative 64; Mismatches 249; Indels 171; Gaps

yy 365 EKPGRRHKCFKAVFGSDALOILHRSHGGERPYKCNVCNGNRPFTTGNLKVHPRHREK 4240
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
yb 407 EKP---FKCSHCDKKFTERSALAKHQHTHNGEKRYKSDSGKETQSNLILHQRITGE 4650
yy 425 YPHVQMMRPHVPEHLDVYITSSGLPIGMSVPEKAEAEATPGGVERKPLVASTALSA 4840
    |||
yb 464 RPY-----4660
yy 485 TESLTLLSTAGTATAGLPAPFKFVLMAKAVEBKNADEMTPGSGSAISGVAESSTA 5430
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
yb 467 ---KCTLCDRR-----FLQNSDLVNHQKYNALP-----LSDPTITAN 5000
yy 544 TLQQLSKMLTSLPSMALLTNHFKSTGSPFLPLCARALGASPSSETSKLQQLVERIKDROGAV 6030
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
yb 501 SPHKCSKCDLTFHSWSTFMKSHLHGKKEKFKQCAEC-----KKGPTOKSDLVKHI 5500
yy 604 AVYSAASGAATTSAPAPSSASSGPNOCVCLFVLSCPARALRYLHGOHNGERPFKCVKG 6630
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
yb 551 RVHTGKK-----PRCKLCKKSPFQNSDLNHNMRHTHGEKFFPCTCD 5930
yy 664 RAFTSTGNLRAHFVVGKASBPAAQNSCPIQOKKFTNAVTLQOHVHRMILGGQIPNGGTAL 7230
    ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

Db      594 KSFTERSALIKH---HRTHTGER-PHKSCVCQKGFIOKSLATKHSRTH----- 637
Oy      724 PEGGGAQWENGSESTVSAGSFPQOQSOQSPSEELSEEEDEEDYDEDSLAG 783
Db      638 -----TGERKYPCTQCGKSFIONDLVKHQRHHTHGERKYHCTECKRRTRESSLVK 688
Oy      784 RGSSESGEKAIIVRAGDSEASGAEEV-----GTVAATAAGKEMDSNEKTTQOSSLP 836
Db      689 HRRTHSGEKYPCPCQCKERTFOSSDLVKHLVHNGENPNPAATAFHEILLRRETLTRSEPD 748
Oy      837 PPPPPDLDPQPMWEGSSGVLGKKEEGKPERSSSPASALU-----PEGEATSVTLVEEL 892
Db      749 PYCTEC-----GKVFQRALLKHLTHKTRKRYPCNECKR-SFQFS 791
Oy      893 SIOEAMKREPESSSRACEVCGAFPQOALEHOKTHPKKGLFTVCYCRGFLERAT 952
Db      792 DLVKHRLTHHTGERPYH--CEPCNKGFIONSDLVKHQHTHTGERP-YTCSQCDKGFIORSA 848
Oy      953 LKKHMLLAHHQVQPF 967
Db      849 LTKHM-RTHTGERKY 862

RESULT 16
ZN91_HUMAN
ID      ZN91_HUMAN          STANDARD;          PRT;      1191 AA.
AC      005481;
DT      01-JUN-1994 (Rel. 29, Created)
DT      01-JUN-1994 (Rel. 29, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Zinc finger protein 91 (Zinc finger protein HNF10) (HPF7).
GN      ZNF91.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_TaxID=9606;
RN      [1]
RS      SEQUENCE FROM N.A.
RA      MEDLINE=93232677; PubMed=8467795;
RA      Bellefroid E.J., Poncelet D.A., Coulle P.G., de Jong P.J.,
RA      Amentiya C.T., Poncelet D.A., Coulle P.G., de Jong P.J.,
RA      Szpirer C., Ward D.C., Martial J.A.;
RT      "Clustered organization of homologous KRAB zinc-finger genes with
RL      enhanced expression in human T lymphoid cells.";
RL      Embo J. 12:1363-1374(1993).
RN      [2]
RS      SEQUENCE OF 15-204 FROM N.A.
RA      MEDLINE=91219421; PubMed=2023909;
RA      Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA      Martial J.A.;
RT      "The evolutionarily conserved Kruppel-associated box domain defines a
RL      subfamily of eukaryotic multifingered proteins.";
RL      Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
CC      -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC      -!- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
CC      FINGER PROTEINS.
CC      -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC      -!- CAUTION: THE SEQUENCE FROM POSITION 1159 TO THE C-TERMINAL IS
CC      DERIVED FROM THE TRANSLATION OF AN ALU REPEAT.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; L11672; AAA59469.1; -
DR      EMBL; M61871; AAA58672.1; ALT_SEQ.
DR      PIR; F39384; F39384.
DR      PIR; S35305; S35305.
DR      HSSP; P08047; ISP2.

```

```
DR   Genew: HGNC:13166; ZNF91.
DR   MIM: 603971.
DR   InterPro: IPR001909; KRAB.
DR   InterPro: IPR000822; ZnF_C2H2.
DR   Pfam: PF00096; ZF-C2H2; 34.
DR   Pfam: PF01352; KRAB; 1.
DR   PRINTS: PR00048; ZINC_FINGER.
DR   PRODOM: PD000003; ZnF_C2H2; 20.
DR   SMART: SM00349; KRAB; 1.
DR   SMART: SM00355; ZnF_C2H2; 33.
DR   PROSITE: PS50805; KRAB; 1.
DR   PROSITE: PS00028; ZINC_FINGER_C2H2_1; 31.
DR   PROSITE: PS50157; ZINC_FINGER_C2H2_2; 35.
DR   Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
KM   Nuclear protein; Repeat.
FT   DOMAIN 13 84 KRAB.
FT   ZN_FING 154 1156 ZINC_FINGERS.
FT   ZN_FING 154 1156 C2H2-TYPE (DEGENERATE).
FT   ZN_FING 179 200 C2H2-TYPE.
FT   ZN_FING 208 232 C2H2-TYPE.
FT   ZN_FING 238 260 C2H2-TYPE.
FT   ZN_FING 266 288 C2H2-TYPE.
FT   ZN_FING 294 316 C2H2-TYPE.
FT   ZN_FING 322 344 C2H2-TYPE.
FT   ZN_FING 350 372 C2H2-TYPE.
FT   ZN_FING 378 400 C2H2-TYPE.
FT   ZN_FING 406 428 C2H2-TYPE.
FT   ZN_FING 434 456 C2H2-TYPE.
FT   ZN_FING 462 484 C2H2-TYPE.
FT   ZN_FING 490 512 C2H2-TYPE.
FT   ZN_FING 518 540 C2H2-TYPE.
FT   ZN_FING 546 568 C2H2-TYPE.
FT   ZN_FING 574 596 C2H2-TYPE.
FT   ZN_FING 602 624 C2H2-TYPE.
FT   ZN_FING 630 652 C2H2-TYPE.
FT   ZN_FING 658 680 C2H2-TYPE.
FT   ZN_FING 708 736 C2H2-TYPE.
FT   ZN_FING 714 742 C2H2-TYPE.
FT   ZN_FING 770 792 C2H2-TYPE.
FT   ZN_FING 820 848 C2H2-TYPE.
FT   ZN_FING 854 876 C2H2-TYPE.
FT   ZN_FING 885 904 C2H2-TYPE.
FT   ZN_FING 932 960 C2H2-TYPE.
FT   ZN_FING 966 988 C2H2-TYPE.
FT   ZN_FING 994 1016 C2H2-TYPE.
FT   ZN_FING 1022 1044 C2H2-TYPE.
FT   ZN_FING 1050 1072 C2H2-TYPE.
FT   ZN_FING 1078 1100 C2H2-TYPE.
FT   ZN_FING 1106 1128 C2H2-TYPE.
FT   ZN_FING 1134 1156 C2H2-TYPE.
SO   SEQUENCE 1191 AA; 137225 MW; 581056b1b8716d CRC64;

Query Match 5.6%; Score 293.5; DB 1: Length 1191;
Best Local Similarity 22.5%; Pred. NO. 3.8e-06;
Matches 141; Conservative 75; Mismatches 239; Indels 173; Gaps 24;
```

```
Db 757 LTFHKRIHTREKPKCKECGKAFIMSSSTLTTRHKRIHTGKEPKYCKCECGKAFRSSTLTTRKH 816
Oy 589 KIQQLVEKIDROGAVAVTSAASGAPTTSAAPSSASSGPNOCVLCRLVSCPRALRLHY 648
Db 817 KTIHTGKEK-----PYCKCECGKAFKSSALAKHK 845
Oy 649 GQHGERPPKCKYCGRAFSTRGNLRAHFVGHKASPARAONSCPIQCKKFTNAVTIQQHV 708
Db 846 ITHAGKLIKCECGKAFNOSNLTTRHKIHTKEPKSSEE----CDKAFIMSTLTTRHK 901
Oy 709 RMHLGQIIPNGTALPEBGGAQENGSEOSTVSGASFPPQOQSOOPSPPELSEEBEED 768
Db 902 RIH-----TRKPKYCKCECGKAFS-----QPSHLTTRHKRMHTGKEPKYCK 940
Oy 769 EEEEDVTDSDSLAGRSSESGEKAISVRGDSSEASGAEEVGYVAATPAGKEMNSE- 827
Db 941 EECGKAFSOSTLTTRHKIHTGKEPKYCK-----CECGKAFKRSSTL-----TEHKIHTGKEK 992
Oy 828 -----KTTQGSLLPPPPPSLDQOPMEQSSGYLGKEGKPERSSPASALTP 879
Db 993 PKYCKCECGKAFSOS-----TLTRHTRMHTGE-----KPKYCKCECGKAFNR 1033
Oy 880 EGEATSVTLVEELSLQDAMRKEPEBSSSRKACEVCGQAFPSQALAEHQTHPRKEGPLFT 939
Db 1034 SSKLTTTHKII-----HTGKEPKYCK-----CECGKAFISSSTLTTRHKRIHTREKP-YK 1079
Oy 940 CVYCRGQFLERATLTKHMLAHNOVPR 967
Db 1080 CEECGKAFSOSTLTTRHKRIHTGKEPKY 1106

RESULT 17
HBL1_CAEEL STANDARD; PRT: 982 AA.
ID HBL1_CAEEL Q9XYD3; Q19389;
AC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hunchback-like protein.
GN HBL-1 OR F13D11.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA MEDLINE=99117349; PubMed=9917360;
RA Fay D.S., Stanley H.M., Han M., Wood W.B.;
RT "A Caenorhabditis elegans homologue of hunchback is required for late
RL stages of development but not early embryonic patterning.";
RL Dev. Biol. 205:240-253(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton L.;
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: REQUIRED FOR LATE STAGES OF DEVELOPMENT.
CC - SUBCELLULAR LOCATION: Nuclear (Probable).
CC - TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN ECTODERMAL CELLS DURING
CC EMBRYONIC AND LARVAL DEVELOPMENT.
CC - SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF097737; AADI6170.1; -
```


Query Match	Best Local Similarity	5.5k; Score 289; DB 1; Length 697;
Matches 110; Conservative	19.7k; Pred. No. 3.5e-06; Mismatches 172; Indels 204; Gaps 14;	
INTERPRO: IPR000637; AT_ho		
INTERPRO: IPR000210; BTF_PO		
INTERPRO: IPR000823; ZnF_C2H		
Pfam: PF000096; zf-C2H2_8.		
Pfam: PF00651; BTF_1.		
Pfam: PF02178; AT_ho		
PRINTS: PR00048; ZINC_FINGER		
PRODOM: PD000003; ZnF_C2H2_		
SMART: SM00384; AT_ho		
SMART: SM00225; BTF_1.		
SMART: SM00355; ZnF_C2H2_8.		
PROSITE: PS50097; BTF_1.		
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.		
PROSITE: PS0157; ZINC_FINGER_C2H2_2; 8.		
Hypothetical protein; Transcription regulation; DNA-binding;		
Zinc-finger; Metal-binding; Nuclear protein; Repeat.		
DOMAIN 10 133 BTF_1		
DOMAIN 294 512 ZINC FINGERS.		
ZN_FING 294 516 ZINC-TYPE.		
ZN_FING 322 344 C2H2-TYPE.		
ZN_FING 350 372 C2H2-TYPE.		
ZN_FING 378 400 C2H2-TYPE.		
ZN_FING 406 428 C2H2-TYPE.		
ZN_FING 434 456 C2H2-TYPE.		
ZN_FING 462 484 C2H2-TYPE.		
ZN_FING 490 512 C2H2-TYPE.		
SEQUENCE 697 AA; 78292 MW; F2BDD33C144626544 CRC64;		

QY	659	CKVCGRAESTGKINLRAHPVGHKASPAARAONSCPICQKKTFTNAVTLQOHWRMHLGGQIPN	718
		! : ! ! ! ! ! : ! : ! ! ! ! ! : ! ! ! ! ! : ! ! ! ! ! : ! ! ! ! ! : ! ! ! ! !	
Db	464	CGIOGKSFSDSSAKRRICILH-----TKKKPFSCPECNLOFARLDNLKALHKIKH-----	512
QY	719	GGTALPBGCGAAOENGSEOSTVSCAGSF-----PQOOS-----	754
Db	513	-----SKEKHADASSISSSSMTVEYRNLTLOLPOLSTSGEOETQLVATDSVN	566
QY	752	-----QQPSPEELISEEHEDEEEDVD	777
Db	563	INFMGPSSGIGISIVTAESSNMFTDAQANTLLTLQOPEQOQNLILLSAQOEQTEHIOQLNM	622
QY	778	EDSLAGSGSESGEKALSYVGDSSEAGAEVEYGTAAATAGKEMDSNKKTTQOSSLP	837
Db	623	IESOMG-PSQTEPRVHTVTLKETETELHAAHQEQTEELHLATSTSDPQHILQLTDEPG--P	679
QY	838	PPPPSDLDQPPQPMEGSS	855
Db	680	PPPTHHVQPTPLQEQSS	697


```
DR EMBL: AF277623; AAK61306.1; -
DR EMBL: AB046831; BAB13437.1; -
DR EMBL: AK024442; BAB15732.1; -
DR HSSP: P08047; 1SP2
DR InterPro: IPR001909; KRAB
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 20.
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PR00048; ZINCFINGER.
DR PRODOM: PD000003; znf_C2H2; 2.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; znf_C2H2; 20.
DR PROSITE: PS00805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 20.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 20.
KW Hypothetical protein; transcription regulation; DNA-binding;
   Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT DOMAIN 8 79
FT ZINC_FINGERS.
FT ZN_FING 257 811
FT ZN_FING 257 279 C2H2-TYPE.
FT ZN_FING 285 307 C2H2-TYPE.
FT ZN_FING 313 335 C2H2-TYPE.
FT ZN_FING 341 363 C2H2-TYPE.
FT ZN_FING 369 391 C2H2-TYPE.
FT ZN_FING 397 419 C2H2-TYPE.
FT ZN_FING 425 447 C2H2-TYPE.
FT ZN_FING 453 475 C2H2-TYPE.
FT ZN_FING 481 503 C2H2-TYPE.
FT ZN_FING 509 531 C2H2-TYPE.
FT ZN_FING 537 559 C2H2-TYPE.
FT ZN_FING 565 587 C2H2-TYPE.
FT ZN_FING 593 615 C2H2-TYPE.
FT ZN_FING 621 643 C2H2-TYPE.
FT ZN_FING 649 671 C2H2-TYPE.
FT ZN_FING 677 699 C2H2-TYPE.
FT ZN_FING 705 727 C2H2-TYPE.
FT ZN_FING 733 755 C2H2-TYPE.
FT ZN_FING 761 783 C2H2-TYPE.
FT ZN_FING 789 811 C2H2-TYPE.
SQ SEQUENCE 818 AA; 94171 MW; 759FDB6338A0D28D CRC64;

Query Match 5.5%; Score 288; DB 1; Length 818;
Best Local Similarity 20.7%; Pred. No. 4.6e-06;
Matches 131; Conservative 66; Mismatches 249; Indels 188; Gaps 18;

368 GGNHKKCFCAKVPSSDSKLOLHLSHTGERPYKCNVCNGRFTTRGNLKVPHRRKPYH 427
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
254 GKPYKCNCKCAKFTQNSMLTSHRRIHSEKPYKSCGKFTFVRSMLTIHQVHTGKPY 313
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
428 VQNNPHVPEHLVDVITSSGLPYGMSVPEKAEENATPGGVERKPLVASTALATES 487
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
314 KCECCGVFRRHNSLTLAHRRIHTGK--PYKCN----- 345
488 LTLTTSAGTATAGLPANFKVFLMKAVEPKNKADENTPGSEGSATSGVAESSTATLM- 546
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
346 ----- 363
547 -----QLSKMTSLSPSMLLTNHRK-STGSPFLPL--CARALGASPSSTKLOOLV 594
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
364 TGEKPFKCNCGKLFTON---SHLISHRNRIHTGKPYKCNCKKAF--SVRSSLAHQTI 418
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
595 EKIDROGAVAVTASASGAPTTASAPASSASSGPNOCVICRLVLSCPRALRLHVGONGGE 654
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
419 HTGEEK-----PYKCNCKGKVFRRNSYLGRRHRRVHTGE 450
655 RPKCKVCYGAFTSGNLRAHFVGHKASPARAONSCPIQCKKETAVALTQOHVHMLGG 714
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
451 KPYKCNCKGKAFSMHNSLNTAHOVHTGKPFK---CNCKSVKFTQNSQLANHRRIHTG- 505
715 QIRNGRGALPBGGAADNGSEQSTVSGAGSPFOOQSOPSPPEELSEEEDEEED 774
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
506 -----EKPYKCNCKGKAFSV-----RSSLITTHQAIHSEKPYKCIKCGKS 545
```

```
OY 775 VTDEDSLAGRSSESGEKAISVRGDSSEASGAEEVGTAAAT----- 818
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 546 FTQKSHLRSHHGHIHSGEKPYK-----NECKGVFAQTSQALAHRRVHTGKPYK 594
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 819 ---AGKEMDSNEKTTQGSLLPPPPDDSLDQPMEDQSSGVLGK--EEGKPERSSSP 873
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 595 CNDCGRAFSRDSRLTFHQAIHTGKPYKCEGKVFRRNSYLTATHRIHTGKPYKCNCK 654
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 874 ASALTPEGEATSVTLVEELSLDEAKRKFESESSRKACEVCGCAFPSQALFEHQTHPK 933
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 655 GKAFSMHNSLTHKVI-----HTGKPYK--CNQCGKVFQNSHLANHQRTHTTG 701
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 934 EGPLFTCVFCROGFLEATLKKMLLAHHVOPE 967
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 702 EKP-YRCNECKGKAFSVKSSLTTHQAI-HTGKPY 733

RESULT 21
ZFHL DROME STANDARD; PRT: 1060 AA.
ID ZFHL DROME
AC P2816;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc-finger protein 1 (zinc-finger homeodomain protein 1).
GN ZFH-1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92001539; PubMed=1680376;
RA Fortini M.E., Lai Z., Rubin G.M.;
RT "The Drosophila zfh-1 and zfh-2 genes encode novel proteins
   containing both zinc-finger and homeodomain motifs.";
RL Mech. Dev. 34:113-122(1991).
CC -1- FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL
   NERVOUS SYSTEM, EMBRYONIC MESODERM AND ADULT MUSCULATURE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: MESODERM AND MESODERMALLY-DERIVED STRUCTURES
   IN THE EMBRYO INCLUDING THE DORSAL VESSEL, SUPPORT CELLS OF THE
   GONADS, AND SEGMENT-SPECIFIC ARRAYS OF ADULT MUSCLE PRECURSOR.
CC ALSO IDENTIFIED IN MOTOR NEURONS OF DEVELOPING CNS.
CC -1- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
   FINGER/HOMEODOMAIN PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 HOMEODOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@sib-sib.ch).
CC
DR EMBL: M63449; AAA29050.1; -
DR PIR: S27816; S27816.
DR PIR: S33641; S33641.
DR HSSP: P08153; 1ZFD.
DR TRANSFAC: T00919; -
DR FlyBase: FBgn0004606; zfh1.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF00046; homeobox; 1.
DR Pfam: PF00096; zf-C2H2; 9.
DR PRINTS: PR00048; ZINCFINGER.
DR PRODOM: PD000003; znf_C2H2; 2.
DR PRODOM: PD000010; Homeobox; 1.
DR SMART: SM00389; HOX; 1.
DR SMART: SM00355; znf_C2H2; 9.
DR PROSITE: PS00027; HOMEODOMAIN_1; 1.
```



```
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PR00048; ZINC_FINGER.
DR ProDom: PD000003; ZnF_C2H2; 17.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; ZnF_C2H2; 33.
DR PROSITE: PS00805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 33.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 34.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 7 75
FT ZINC_FINGERS.
FT ZINC_FINGER (DEGENERATE).
FT ZN_FING 201 1147
FT ZN_FING 201 1223
FT ZN_FING 229 251
FT ZN_FING 257 279
FT ZN_FING 285 307
FT ZN_FING 313 335
FT ZN_FING 341 363
FT ZN_FING 369 391
FT ZN_FING 397 419
FT ZN_FING 425 447
FT ZN_FING 453 475
FT ZN_FING 481 503
FT ZN_FING 509 531
FT ZN_FING 537 559
FT ZN_FING 565 587
FT ZN_FING 593 615
FT ZN_FING 621 643
FT ZN_FING 649 671
FT ZN_FING 677 699
FT ZN_FING 705 727
FT ZN_FING 733 755
FT ZN_FING 761 783
FT ZN_FING 789 811
FT ZN_FING 817 839
FT ZN_FING 845 867
FT ZN_FING 873 895
FT ZN_FING 901 923
FT ZN_FING 929 951
FT ZN_FING 957 979
FT ZN_FING 985 1007
FT ZN_FING 1013 1035
FT ZN_FING 1041 1063
FT ZN_FING 1069 1091
FT ZN_FING 1097 1119
FT ZN_FING 1125 1147
SQ SEQUENCE 1167 AA; 134352 MW; E2184DF3BD035E9 CRC64;
Query Match 5.3%; Score 278.5; DB 1; Length 1167;
Best Local Similarity 21.9%; Pred. No. 1.7e-05;
Matches 143; Conservative 76; Mismatches 266; Indels 167; Gaps 24;
```

```
DR 703 KPYKCECGKTFESKVSILTTHKAIHAGEKPYKCECGKAFNSKFSILTTHKHYHTEKPYK 762
QY 687 AQNSCPICQKKFTNAVTLQOHVHMLGGQIPNGGTALPREGGAAGENGSDQSTVSGAGSF 746
DB 763 ---CEECGKAYKWPSTLSYHKHIHTG----- 785
QY 747 PQQSOQPSPEELSEEEDEEEDVTDDESLAGRSGSGGEKA---ISVRGDSSEA 803
DB 786 ----EKPYKCECGKSGMSILTKHEVYIHGEKPYKCECG--KAFSLIVSKHKKT 838
QY 804 SGAEVEGYVAAATAGKEMDSNEKTTQSSLP PPPPSLD P Q P M E G S S G V L G K E 862
DB 839 HAGEK----FYKCEACGKAYNTFSILTTHKHYHTEKPYKCECGKAFNMSSNLMHKKI 894
QY 863 -EGKP-----ERSSSPASALTPGEGATSVYLVELSLQEAHMKRKEGSSSRKACEVCG 915
DB 895 HTGETPYKCECGCDKAFSPSSILT-EHKAT-----HAGEKPYK--CECG 935
QY 916 QAFPSOALAEHOKTHPKEGPLFTGCVGROGFLERATLKKMLLAHQVOFP 967
DB 936 KAFSPSRILTEHKATHAGEEP-YKCECGKAFNMSSNLMHKKI-HTEKPY 985

RESULT 23
Y222.HUMAN STANDARD: PRT: 1163 AA.
AC 092618;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein KIAA0222.
GN KIAA0222.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
CC -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: D86975; BAAL3211.1; -.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF00006; Zf_C2H2; 10.
DR PRINTS: PR00048; ZINC_FINGER.
DR ProDom: PD000003; ZnF_C2H2; 1.
DR SMART: SM00355; ZnF_C2H2; 10.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT DOMAIN 34 1120
FT ZN_FING 34 56
FT ZN_FING 62 84
FT ZN_FING C2H2-TYPE (ATYPICAL).
```

FT	2N_FING	174	197	C2H2-TYPE.	
FT	2N_FING	200	223	C2H2-TYPE (ATYPICAL).	
FT	2N_FING	248	270	C2H2-TYPE.	
FT	2N_FING	276	298	C2H2-TYPE.	
FT	2N_FING	335	357	C2H2-TYPE.	
FT	2N_FING	515	537	C2H2-TYPE.	
FT	2N_FING	760	783	C2H2-TYPE (ATYPICAL).	
FT	2N_FING	1098	1120	C2H2-TYPE.	
SO	SEQUENCE	1163 AA;	124288 MM;	9D6DDDED2D70236A4 CRC64;	
Query Match					
Best Local Similarity 5.3%; Score 278; DB 1; Length 1163;					
Matches 161; Conservativity 21.0%; Pred.No. 1.8e-05;					
Matches 161; Conservativity 75; Mismatches 266; Indels 264; Gaps 28;					
OY	362	GPLEKPGGR-----	-HKRCFAKVFVGSLSALQIHLRSHTEGRPYKCNOCNPFET	410	
DB	14	GSPTTRAGRHEVDGKATCTCCICGSPFOSSLSQHMRKHTGKPCYCDHRSQ		73	
OY	411	RCNLAKVHFRHREKYPVHOMNHPREHLIDYVITSSGLPYGKSVPEKAEAEATPGGV		470	
DB	74	KGNLKIHRSHRTGTLQGHPEAGAPLGEMRASGL-----	-DACASP----	117	
OY	471	ERKPLVASTALSTESITLLTSTAGTAPGLPAFNKFLVLMKAVEPKKAD-----	EN	524	
DB	118	-----FKSASACNRL-----	NGASQADGARYLN-----	GASQADSGRVLLRS	155
OY	525	TPPGSGSATSIVNESSATILMOLSKLMTSLPSMALLTNH-FKSTGSFPLICARALGAS		583	
DB	156	SKKGABESA---CAPGEAKAVVOCSECKSOFRKKDLHLVHOAHKPFKCRCLSYA----		208	
OY	584	PSETSKLOOLVEKIDROGAVAVTSAASGAPTTSAAPSSASSGPNQCVLCRLVSCPRA		643	
DB	209	---TLRESLSHIERD---HITAGPGSGEACVENGKPELSGEPCEVCGAQFSQTFW		262	
OY	644	LRHNGOHGEBRPPCKVCYCGAFTSGNIRAFVGH-----	-KASPARAON--	689	
DB	263	LKAHMKHRSFPHGCHICGRFRKEPWFILKNMKAHGPTGSKNRPKSELDPATINNIV		322	
OY	690	-----SCPIQCKFTNAVTLQOHVHMLGGQIPNGGTLPEGGGAQENGSE		736	
DB	323	QEEVYIAGSLIYECACACGNLFTNLIDSLAHNAIHF--RVEASRTAPAEAGAGPSDTK		380	
OY	737	QSTV-----SGAG-SFP-----	-OOQSOQPSPEELSEEEE-----	765	
DB	381	QEPLOCLNLRPSAAGDSCTQAGRRVAELDPVNSYOAWOLATRGVAEPAYELIKYGAMD		440	
OY	766	-----BEDEEE-----	EEDVTDEDSLARG-----	SESGEKATSVRGDSEASG	805
DB	441	EALAGVAFPKDRREVLYVSOEKRRREDAPAAQGP RRKRASGDPAPAPAGHLDP RSAAR		500	
OY	806	AEEEVGTVAATAATAG-----	-KEMDSNEKTTQOSS	834	
DB	501	PNRR-----	AAATTGQCKSSSECFEGCKIFRTYHOMVLSRVRARRRERSDDGRARAR	555	
OY	835	LPPPPPDSDLOPOMBOGS-----	SGVLGKE--EGKPER---	869	
DB	556	CSLSGDSASQPS--SPGSACAADSPGSLADEAEDSGEGGAPAGGQPRRCFS		613	
OY	870	-----	SSPASALTPBEATSVTLVEELSLDEAMKE-----	901	
DB	614	EELYTSELSSGDOSHKMGNDNSERDTEGSKAGIAASVSLIENSSRSTSRQEOHFRSMDL		673	
OY	902	-----	PGESSRKACEVCGQAFPSQALAEHQHTPK	934	
DB	674	KPAFHFKQEVVPFGD-----	-GVEFBSSTG-ABEQOTGHPAE	708	
RESULT 24					
ID	2133_HUMAN	STANDARD;	PRT;	654 AA.	
AC	P52736; Q9H443; Q9BUV2;				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				

DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Zinc finger protein 133.
GN	ZNF133.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RC	SEQUENCE FROM N.A.
RP	TISSUE=Insulinoma;
RX	MEDLINE=95377390; PubMed=7649249;
RA	Vissing H., Meyer W.-K., Aagaard L., Tommerup N., Thiesen H.-J.;
RT	"Repression of transcriptional activity by heterologous KRAB domains
RT	present in zinc finger proteins.";
RL	FEBS Lett. 369:153-157(1995).
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=21638749; PubMed=11780052;
RA	Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA	Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA	Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA	Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA	Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA	Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA	Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA	Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA	Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA	Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA	Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA	Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA	Lhevassialho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA	Marsh V.L., Martin S.L., McConachie L.J., McIay K., McMurray A.A.,
RA	Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA	Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.T.,
RA	Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA	Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA	Skue C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA	Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA	Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA	Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA	Wilmink L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA	Rogers J.;
RT	"The DNA sequence and comparative analysis of human chromosome 20.";
RL	Nature 414:865-871(2001).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Lung;
RA	Strausberg R.;
RL	Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
CC	- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION AS A
CC	REPRESSOR.
CC	- SUBCELLULAR LOCATION: Nuclear (Potential).
CC	- TISSUE SPECIFICITY: SEEMS UBIOUITOUS. SEEN IN THE HEART, BRAIN,
CC	PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, AND PANCREAS.
CC	- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC	FINGER PROTEINS.
CC	- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; U09366; AAC50260.1; -
DR	EMBL; AL049646; CAC15508.1; -
DR	EMBL; BC001887; AA01087.1; -
DR	HSSP; P08045; 1ZNF.
DR	TRANSFAC; T04992; -
DR	GeneW; HGNC:12917; ZNF133.

DR EMBL: AF317549; AAG59817.1; -
DR EMBL: AF385187; AAK69307.1; -
DR EMBL: X78926; CAAS5526.1; -
DR HSSP: P08047; 1SP2.
DR Genew: HGNC:13061; ZNF268.
DR MIM: 604753; -
DR MIM: 606024; -
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 48.
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PRO0046; ZINCFINGER.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; Znf_C2H2; 24.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 24.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 24.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 24.
DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat; Alternative splicing.
FT DOMAIN 81 152 KRAB.
FT ZINC_FINGERS.
FT ZN_FING 276 942 ZINC_FINGERS.
FT ZN_FING 276 298 C2H2-TYPE.
FT ZN_FING 304 326 C2H2-TYPE.
FT ZN_FING 332 354 C2H2-TYPE.
FT ZN_FING 360 382 C2H2-TYPE.
FT ZN_FING 388 410 C2H2-TYPE.
FT ZN_FING 416 438 C2H2-TYPE.
FT ZN_FING 444 466 C2H2-TYPE.
FT ZN_FING 472 494 C2H2-TYPE.
FT ZN_FING 500 522 C2H2-TYPE.
FT ZN_FING 528 550 C2H2-TYPE.
FT ZN_FING 556 578 C2H2-TYPE.
FT ZN_FING 584 606 C2H2-TYPE.
FT ZN_FING 612 634 C2H2-TYPE.
FT ZN_FING 640 662 C2H2-TYPE.
FT ZN_FING 668 690 C2H2-TYPE.
FT ZN_FING 718 746 C2H2-TYPE.
FT ZN_FING 752 774 C2H2-TYPE.
FT ZN_FING 802 830 C2H2-TYPE.
FT ZN_FING 836 858 C2H2-TYPE.
FT ZN_FING 864 886 C2H2-TYPE.
FT ZN_FING 892 914 C2H2-TYPE.
FT ZN_FING 918 942 C2H2-TYPE.
FT ZN_FING 920 942 C2H2-TYPE.
FT VARSPLIC 1 161 MISSING (IN ISOFORM B).
FT CONFLICT 860 860 R -> T (IN REF. 2).
FT SEQUENCE 947 AA: 108373 MW: AC76F4824F4BE1A0 CRC64;
Query Match 5.28; Score 276.5; DB 1; Length 947;
Best Local Similarity 21.6%; Pred. No. 1.7e-05;
Matches 133; Conservative 57; Mismatches 248; Indels 179; Gaps 20;
365 EKPGRHKKCFCKKVFSSDALQIHRSHTGERPKCNVCGRNFTTRGNLKVHFRHREK 424
DB 385 QRP---YVCEGCGAFKLSQLIHERIHTEKPYECNCGAKFNKSNLMVHQRTTGE 441
425 YRHWQNNPVRPHLDVITSSGL-----PY-----GMSYPPEKAEAEATPGGVER 472
DB 442 KRYVCDGCKAFPKFSQLIYQVGIHTGVKPRGCCICGCGFSLKSQLIYHQRSHTG----M 497
473 KPLVASTTALSTATESLTLLSTASGTATAPAGLPAPNKFVLMKAVEPRKNKADENTPPGSEGS 532
DB 498 KRYVCEGCG-KAERKSKSYLLIHTRTHTGEKLEHCNNC-----GK 535
533 ATSGVASSATITLMQSLKMTSLPSMALLTNHFKSTGSPPLP--CARALGASPSSTSL 590
DB 536 ASESFKSQ-----LIHQRIHTGENPYECHECGKAF-----SRK 568
591 QOLVERIDRGAAVAVTSAAGAPTTSAAPAPSSASSGPNOCYICLVLSCPRALRLHYGQ 650
DB 569 YDLISH-----QRTNHAQEKPYECIDCGKAFGLKSQLIHQRT 605

QY 651 HGERPEKCKVCGRAFTSRGNLRAHVGHKASPARAONSCPICQKKTNAVTLQOHVRM 710
DB 606 HTGKRPFCSCQKAFNFKNSL-----IYHQRTHTGKRPYSCNCGAFTFKLSQLIYHKG 661
QY 711 HLGGQIPNGGTALPEGGGAOENGSEOSTVSGASFPOQSOQPSPEELSEEEDEDEE 770
DB 662 HTG-----VKPYCCSQCAKTFSLKSQ-----IVHORSHGVKPYCCSE----- 700
QY 771 EEEVDVEDSLAGSGSESGEKAIIVRGDSEASGAEEVGTAAATACKEDMSNKT 830
DB 701 -----CGKAFNKSQYLLIHMRTHTGKPHCRE-----CGKSSFNLSQLI 740
QY 831 QQSLEPPPPPSLDQPOPMEGSGSSGLVGKDEGKPKERSSPASALTPREGATSVTLVE 890
DB 741 VHQRI-----HTGENPYECSECGKAFNRNDQLS----- 769
QY 891 ELISQEMRKREPPGSSSSKACEVCGAOPSOALAEHQTHREKPLFTVFCROGLER 950
DB 770 -----HQRTHAGKRP--YGCSECGKAFSSKSYLLIHMRTHSGEKP--YECNECGKAFIWK 820
QY 951 ATLKKHMLLAHVOQPF 967
DB 821 SLIIVHE-RTHAGVNPY 836
RESULT 26
EVL_MOUSE
ID EVIL_MOUSE STANDARD; PRT: 1042 AA.
AC P14404;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ecotropic virus integration 1 site protein.
GN EVIL OR EVI-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88311086; PubMed=2842066;
RA Morishita K., Parker D.S., Mucenski M.L., Jenkins N.A., Copeland N.G.,
RA Thle J.N.;
RT "Retroviral activation of a novel gene encoding a zinc finger protein
in IL-3-dependent myeloid leukemia cell lines".
RL Cell 54:831-840(1988).
CC -!- SUBUNIT: MAY INTERACT WITH CBP1 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-11 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M21829; AAA40581.1; ALT_INTT.
DR PIR: A31591; A31591.
DR HSSP: P08047; 1SP2.
DR TRANSFAC: T00273; -.
DR MGD: MGI:95457; Ev11.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 10.
DR PRINTS: PRO0048; ZINCFINGER.
DR ProDom: PD000003; Znf_C2H2; 4.
DR SMART: SM00355; Znf_C2H2; 10.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 10.
KW Zinc-finger; Metal-binding; DNA-binding; Developmental protein;
KW Repeat; Proto-oncogene; Chromosomal translocation.

```
FT DOMAIN 21 239 ZINC-FINGERS 1.
FT DOMAIN 421 434 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 724 803 ZINC-FINGERS 2.
FT DOMAIN 877 928 ASP/GLU-RICH (ACIDIC).
FT ZN_FING 21 44 C2H2-TYPE.
FT ZN_FING 75 97 C2H2-TYPE.
FT ZN_FING 103 125 C2H2-TYPE.
FT ZN_FING 131 154 C2H2-TYPE.
FT ZN_FING 160 182 C2H2-TYPE.
FT ZN_FING 188 210 C2H2-TYPE.
FT ZN_FING 217 239 C2H2-TYPE.
FT ZN_FING 724 746 C2H2-TYPE.
FT ZN_FING 752 775 C2H2-TYPE.
FT ZN_FING 781 803 C2H2-TYPE.
FT SITE 553 557 CTBP-BINDING MOTIF 1 (BY SIMILARITY).
FT SITE 584 588 CTBP-BINDING MOTIF 2 (BY SIMILARITY).
SQ SEQUENCE 1042 AA: 116847 MW: 8DEDF164F536D2FE CRC64:

Query Match 5.2%; Score 276; DB 1; Length 1042;
Best Local Similarity 18.7%; Pred. No. 2e-05;
Matches 176; Conservative 87; Mismatches 245; Indels 432; Gaps 34;

OY 368 GGRHNRKFRCAKVGSDSALQIHLDS-HTGERPKVCNVCGRFTTGNLKVHFRHREKYP 426
D 128 GKHYECENCAKVFDPNLRHRSQHVGAFAHACBCECKTFATSSGLKQHKHHSVKP 187
OY 427 -----HYQMN----- 431
D 188 FICEVCHKSTQFSNLCRHKRMHADCRTQIKCKDCGOMFSTTSLNKHRRFCEGKNHFAA 247
OY 432 -----PHRPEHLDYVITSSGLRPGMSVPEPKA---EEEA 464
D 248 GGFEGGIGISLPGTPAMKTSVNVNMSHANPGLADYGTNNH-PAGLTFPTAPGFSFSPGL 306
OY 465 TPEGGVKRPVLAST--TALSAFE-----SLTLSTAGATAPGCPAFNKFVLM-- 512
D 307 FPGGLYHRPPLIPASRPVKGLSTFEOSNKCGSPRLTHPQILPATQDILKALSKHPYGDN 366
OY 513 KAVE--PKNADE-----NTPRGSE-----GSAISGAESSTAILMQ 547
D 367 KPELLPERSSEEPLEKISDQSSSDLDVSTPSGSDLETTSGSDLESDEKKECKE 426
OY 548 LSKL-----MTSLPMSALTN-----HFK 566
D 427 NGMFMKDKVPLQGLASTNNKEHNHNSVFSASVEBQASVSGAVNDSIKALASTAEKTFG 486
OY 567 STG-----SFPLP-----LCARALGASPSSTKIQ 591
D 487 STGLVGLQDKKVGALPYRSMFPLPFPPAFSGSMYPPRDLRSLPLKMEPOSPEVKLIQ 546
OY 592 Q-----LVERIDOGAVAVTSAAGAPTTSAPAP-----SSASSG----- 627
D 547 KGSSESPFDLTKRKDEKPLTSGPSPGTPATISODQPLDLSMGSRGASGTLKLEPRKN 606
OY 628 ----- 627
D 607 HVGEEKKGSNMDTSPSSGSLQHARPTPFPMPIYVEKRLTDPLEALEKYLPSRPF 666
OY 628 -----PN-----QC 631
D 667 LHPHOMSAIENMAEKLESFALKPEASELLQVPSMFSFAPRNTPLPENLIRKKEKERYTC 726
OY 727 RYCGKTI--PRSANLTRLRTHHTGEOPYRCKYCDRFSISSNQRHNRHNKKEPF-- 781
OY 688 QNSCPLCOKFTTNAVTLQOAHVMMHGQIPNGGTALP---ECGGAA----- 730
D 782 --KCHLDCRCGQGTNIDRHLKKHNGMNSGTATSSPHSELASAGALLDKEADVFTPEIR 839
OY 731 -----QENGSE-----OSTVSGAGSFPOQOQOOPSPDEELISEEEDDE---BEEDVT 776
D 840 NFGNSNHGSGSPRNMERMNGS-HFKDKKALATSONSDLLDDEEVEDEVLLDEDEDD-- 896
```

```
OY 777 DEDSLAGRSESGGEKAISVRGDSEASGAEVEVTVAAATAG----- 820
D 897 --NDIPKPKRELG-----VTRUDEELPEDDYEEAGLEMSCKSPVKRYKREDYKSGLSAL 950
OY 821 -----KEMDSNEKT-TQSSSLPPPPPPPSLDQPPMEQSSGVLGKEEGKP 867
D 951 DHTRHFTDSLKRMEENQYTDALSLSSISSHVPELKTQLHKK-----SKQAVAM 1002
OY 868 ERSSSPASALTPEGEATSVTLVEELSLQELAMRKREPESSS 907
D 1003 MLSLSDKDSLHPTSHSSS-----NWMHSMARAAESSA 1035

RESULT 27
ID ZN43_HUMAN STANDARD; PRT; 803 AA.
AC P17038; P28160; Q96DG1;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein 43 (Zinc protein HTF6) (Zinc finger protein KOX27).
GN ZNF43 OR ZNF39 OR KOX27.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-T-cell;
RX MEDLINE=91279444; PubMed=1711675;
RA Lovering R., Trowsdale J.;
RT "A gene encoding 22 highly related zinc fingers is expressed in
RT lymphoid cell lines.";
RL Nucleic Acids Res. 19:2921-2927 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Lymph;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 38-190 FROM N.A.
RX MEDLINE=91219421; PubMed=2023909;
RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA Martial J.A.;
RT "The evolutionarily conserved Kruppel-associated box domain defines a
RT subfamily of eukaryotic multifingered proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612 (1991).
RN [4]
RP SEQUENCE OF 476-531 FROM N.A.
RC TISSUE-Lymphoid;
RX MEDLINE=91145339; PubMed=2288909;
RA Thiesen H.-J.;
RT "Multiple genes encoding zinc finger domains are expressed in human T
RT cells.";
RL New Biol. 2:363-374 (1990).
CC -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: T AND B CELL LINES.
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59244; CAA41932.1; -.
```

```
DR EMBL; BC006528; AAH06528.1; -.
DR EMBL; M61869; AAA58674.1; -.
DR EMBL; X52358; CAA36584.1; -.
DR PIR; S26823; S26823.
DR PIR; D39384; D39384.
DR PIR; S10416; S10416.
DR HSSP; P08048; 7ZNF.
DR TRANSEAC; T04986; -.
DR Genew; HGNC:13109; ZNF43.
DR MIM; 603972; -.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR008422; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 21.
DR Pfam; PF01352; KRAB; 1.
DR PRINTS; PR00048; ZINCINGER.
DR PRODOM; PD00003; Znf_C2H2; 16.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 21.
DR PROSITE; PS50805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 22.
KW Transcription regulation; DNA-binding; zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 1 69 KRAB.
FT 168 778 ZINC_FINGERS.
FT ZN_FING 168 190 C2H2-TYPE.
FT ZN_FING 196 218 C2H2-TYPE (DEGENERATE).
FT ZN_FING 224 246 C2H2-TYPE (DEGENERATE).
FT ZN_FING 252 274 C2H2-TYPE (DEGENERATE).
FT ZN_FING 280 308 C2H2-TYPE.
FT ZN_FING 308 330 C2H2-TYPE.
FT ZN_FING 336 358 C2H2-TYPE.
FT ZN_FING 364 386 C2H2-TYPE.
FT ZN_FING 392 414 C2H2-TYPE.
FT ZN_FING 420 442 C2H2-TYPE.
FT ZN_FING 448 470 C2H2-TYPE.
FT ZN_FING 476 498 C2H2-TYPE.
FT ZN_FING 504 526 C2H2-TYPE.
FT ZN_FING 532 554 C2H2-TYPE.
FT ZN_FING 560 582 C2H2-TYPE.
FT ZN_FING 588 610 C2H2-TYPE.
FT ZN_FING 616 638 C2H2-TYPE.
FT ZN_FING 644 666 C2H2-TYPE.
FT ZN_FING 672 694 C2H2-TYPE.
FT ZN_FING 700 722 C2H2-TYPE.
FT ZN_FING 728 750 C2H2-TYPE.
FT ZN_FING 756 778 C2H2-TYPE.
FT CONFLICT 181 181 P -> S (IN REF. 3).
FT CONFLICT 712 712 P -> S (IN REF. 2).
FT SEQUENCE 803 AA; 93487 MW; 88C41B44DFDE2533 CRC64;

Query Match 5.2%; Score 275; DB 1; Length 803;
Best Local Similarity 20.1%; Conserved 59; Mismatches 194; Indels 282; Gaps 22;
Matches 135;

347 KPRNGSGELSYGVMPLEKPGGRHKCFCAKVFEGSDSALQIHLRSHTGERPKVCNCGN 406
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
287 KAFNGSNLTLEHKIHTGEKP---YKCECGKAFWMPSTLTTHKRIHTGEKPYTCECGK 343
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
407 RPTTRGNLKVHFRH-REKTPHYVOMNPHVPEHLDYITSSGLPYGMSVPPEKAEERAAAT 465
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
344 AENQFSNLTTHKRIHTGEKP-----YKCECG----- 370
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
466 PGGGVARRKLVATTTALSAESTLTLSTAGTATAPGLPAFKFVLMKAVERKNKADENT 525
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
371 -----EASRSNLT----- 390
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
526 PGSE--GSAISGVASSTATIMQSLKMTSLPSW-ALITNHFKSTGSPPLPCARALGA 582
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
391 PKCECGKAF-----KSSKITEH----- 410
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
583 SPSESTKIQQLVLEKIDRQGAVAVTSAASGADPTSPADPSSASSSGPNQVCICRLVISCPR 642
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 411 -----KLTHTEK-----PYKCECGKAFNMP 433
QY ALRHYGQHGGERPKFCVGRAPSTRGNLRAHFVGHKASPARAQNSCPICQKFTNAV 702
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
434 TLTKNRIHTGEKPYKCECGKAFNQSNTLT---HKRIHTAEKPYKCECGKASRSS 489
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY TLOQHVAMHLGGQIPNGGTALPEGGAQENGSEOSTVSGASFPQOQSQPSEBELSE 762
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 490 NLTKHKIHT-----EKPKYKCECGK 511
QY EEEEDDEEEEDVTDDESLAGRSESG-----GKALSYVRDSEASG 805
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 512 AFKWSNLTTHKRIHTGEKPYKCECGKAFNHSILTKRIHTGEKPKYC----- 562
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY AEREVG---TVAAATAGKENDSNEK-----TTQOSSLPPEPPDLPDOPMEQ 852
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 563 ---EECGKFTQSSNLTTHKRIHTGEKPYKCECGKAFNQSNTLTTHKRIHT 611
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY GSSGVLGKKEGKPKERSPPASALTPREGATSVTLVEELSLQAMRKPEGESSRRACE 912
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 612 -----GGKPYKCECGKAF---NQSTLTTHKRIHTGEKPYK-----CE 647
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY VCGQAPPSQALAEHQKTHPEKGLFTVCFCRQGLERATLKKHMLAHQVOPF----- 967
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 648 ECGKAFKWSNLTTHKRIHTGEKP-YKCECGKAFKLSLTSTHKII-HTGEKPYKCEK 705
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 968 --APHGPONI 975
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 706 GKAFNRSNL 715

RESULT 28
2226 HUMAN STANDARD; PRT; 803 AA.
ID Q9NT6; Q9NS44; Q96TES;
AC 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE zinc finger protein 226.
GN ZNF226.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
RT Stubbs L.;
RA "Differential expansion of homologous zinc-finger gene families in
RT human chromosome 19q13.2 and mouse chromosome 7.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Kodoyianni V., Ge Y., Severin J., Krummel G.K., Grable L.,
RA Kvilstad E., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.;
RT "Sequence analysis of a 1Mb region in 19q13.2 containing a zinc finger
RT gene cluster.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shitatori A., Sudo H.,
RA Magatsuma M., Hosotani T., Kaku Y., Kodaira H., Kondo H.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
```



```

HSSP; P08047; 1SP2.
InterPro; IPR000822; znf_C2H2.
Pfam; PF00096; zf-C2H2; 15.
SMART; SMO0355; znf_C2H2; 15.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.
PROSITE; PS50157; ZINC_FINGER_C2H2_2; 15.
Zinc-finger; Metal-binding; DNA-binding; Repeat.
NON_TER 1 1
T 2N_FING 250 272 C2H2-TYPE.
T 2N_FING 278 300 C2H2-TYPE.
T 2N_FING 306 328 C2H2-TYPE.
T 2N_FING 334 356 C2H2-TYPE.
T 2N_FING 362 384 C2H2-TYPE.
T 2N_FING 390 412 C2H2-TYPE.
T 2N_FING 418 440 C2H2-TYPE.
T 2N_FING 446 468 C2H2-TYPE.
T 2N_FING 474 496 C2H2-TYPE.
T 2N_FING 618 640 C2H2-TYPE.
T 2N_FING 646 668 C2H2-TYPE.
T 2N_FING 674 696 C2H2-TYPE.
T 2N_FING 702 724 C2H2-TYPE.
T 2N_FING 730 752 C2H2-TYPE.
T 2N_FING 758 780 C2H2-TYPE.
T NON_TER 780 780
SEQUENCE 780 AA; 88648 MW; 4886C5BEFC9B3915 CRC64;

Query Match 5.1%; Score 270; DB 1; Length 780;
Best Local Similarity 18.4%; Pred. No. 2,7e-05;
Matches 163; Conservative 81; Mismatches 308; Indels 332; Gaps 30;

149 PLPPSTAPAPPPPP-----PPPPPGVSGGLNIP 178
144 PTNPISPKKOPPPANGIKEVASCDGROSDCSINPLETOIGDTDIPPIGYNHLIIQ 203
179 LILE-----ELRVLOQROIHOMOTEOICROYLLGLSGTQVAPASPSLPGTGTAS 231
204 AIKDMENTSPHSRFRNHTHTTE--KQFSNHNINLNDLSIHAGKKPPRCSCGCGCFAG 261
232 STKPLP--PLFSPRIKPVQTSKTLASSSSSSSSSGAETPKQAFPHLYPLGSOHPFSAG 289
262 SSELVNHRRTHRVKVF-----SCSOGCKCFSTQTKLKHNRHTH--GEKPPSCSE 310
290 VGRSKRPAPRPALPGSIDQLASPHL-----AFPSTTGLLAOCLGARGCL 337
311 CGK-----CPSTPHVARHOKTHGEGKPPSCSE--CGKCF--ARSS 347
338 EATASPGLLKPKNGSGELSYGEVWPLEKPGRAKRCRCACAVFGSDALQIHLSHTGER 397
348 DVT-----VHRKTHTG-----EKP--YSCSGCKCFTRSSDLNVRHRTHTGK 388
398 PYKCNVCGNRFTTNGMLKYFHRRHREKYPHYQMNPHVPEHLDVYITSSGLPYGMSVPP 457
389 PYSCGHGCKCFSTSSSLNVRHRTHTGKPY----- 418
458 KAEBAATPVGGEVERKRPVASTALSTESTLTLSTAGTATAPGLPAFNKVLMAVER 517
419 ----- 418
518 KNAKADENPPGSGAISGVASSTATIMQLSKLMTSLPSWALLTNHPKSTGSFPLPLCA 577
419 -----SCSECGKCFPTTSEFISHKT----- 439
578 RALGASPSETSKLOOLVEKIDROGAVAVTSAAAGAPTTSAAPSSASSGPNOCVTCILRV 637
440 -----HMEKK-----PPSCVCGCK 454
638 LSCPALRIHYGONGGEERPKCKVCGRAPSTRGNILRAHFGVGHKASPARAAQNSCPIQK 656
455 FSKDTHLKHYHTHTEKCPSCPCGCKCFTHNGSLKYLKTHK-----READFCSKG 506
697 KFTN-AVTLQOHVBMILGGQIPNGGTALPRGGGAOENGSEOSTVSGAGSPFOQSQOPS 755
507 NLTNPEISVEVHP--TNEIKBATSWEBGNQSDYSINSLTEQIQ-----PYTPI 556

```

```

QY 756 PEE-----LSEEEDEDE-----EEDVDTEDSLACGSGSGEKAISVRCDSEFASGA 806
D 557 PIMEXNHLIMQDNKYDYNACHSPLOETDVT-KNALHKRIDIRORTQTLQKDYHRTNTD 615
QY 807 EEEVGVAAATATGKEMDSNEKTTQSSLPRRPPRDSIDQOPMIGSS--GVLGKEEG 864
D 616 KP-----LSCSECGKCFSTYVILARHOKTHGKPFSCSECKYARSDLLNVRHTHTG 670
QY 865 GKPERSSPASALTPGEATSVTLVEELSLQEAARKKEGESSRRACEVCGQAFPSQAL 924
D 671 EKPYSCSECGKCFTRSSD-----FNVHRTHTGKRP--YSCSGCKCFPTSSVL 717
QY 925 EEHQKTHPEKGLFTCVFCRQGLERATLKKHMLIAHQVDF 968
D 718 TSHWRHTHTGKRP-FSCTECGKCFSTETLYKYNH-RTHGKERFS 759

RESULT 30
PRDH_HUMAN STANDARD; PRT; 720 AA.
ID O9H9D4;
DC 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE PR-domain zinc finger protein 17.
GN PRDM17 OR PFM14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RA Du Y., Huang S.;
RT "A family of novel PR-domain (PRDM) genes as candidate tumor
RL suppressors."
RL Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;
RT "NDO human cDNA sequencing project."
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RA SEQUENCE FROM N.A.
RC TISSUE=Lung, and Uterus;
RA Strausberg R.;
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF346626; AAK29075.1; -.
CC EMBL; AK028893; BAB14295.1; -.
CC EMBL; BC013355; AAH13355.1; -.
CC EMBL; BC015708; AAH15708.1; -.
CC HSSP; P08153; 1ZPD.
CC InterPro; IPR000822; znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 10.
CC ProDom; PD000003; znf_C2H2; 1.
CC SMART; SMO0355; znf_C2H2; 10.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.
CC PROSITE; PS50157; ZINC_FINGER_C2H2_2; 10.

```

Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT ZN_FING 353 375 C2H2-TYPE.
FT ZN_FING 381 403 C2H2-TYPE.
FT ZN_FING 409 431 C2H2-TYPE.
FT ZN_FING 437 459 C2H2-TYPE.
FT ZN_FING 468 480 C2H2-TYPE.
FT ZN_FING 496 518 C2H2-TYPE.
FT ZN_FING 524 546 C2H2-TYPE.
FT ZN_FING 551 573 C2H2-TYPE.
FT ZN_FING 579 601 C2H2-TYPE.
FT ZN_FING 607 629 C2H2-TYPE.
SQ SEQUENCE 720 AA: 78438 MW: 103CDA7572D17721 CRC64;
Query Match 5.1%; Score 269.5; DB 1; Length 720;
Best Local Similarity 20.68; Pred. No. 2.6e-05;
Matches 181; Conservative 77; Mismatches 315; Indels 307; Gaps 36;
QY 7 RSSRLG PAG-EPAELGDASEEDHPQVCAKCAQFTDPT-EFLAHQNA----- 53
DB 19 REPRLG DLGWNPSGEGCTGGLKDY-----PEPTRILALKSLPRGLALGSLA 68
QY 54 -----CSTDPPVAVYIGQENPNMNSASSERPEGHNNPDVMDTEHSNPPDSSGSSVP 105
DB 69 KEORLG YMGVGDPLQGLTWG---PLEEESASKEKEGVK-----PROEENTLS 113
QY 106 TDPTNG EERGSESSGHFLVAATGTAAGG-----GLLASKLGLATPL 150
DB 114 LGP-WGDVACAEQSSGWTSLVQKRLSEGVAVRISERLHLYQVQLVPLGSELLMPQ 172
QY 151 PEEPPAPRRPPRRPPVGVSGHINIPRLIEELRYLQORIQHOMQTRQICQVLLLG 210
DB 173 PSESGS -----LTQPGDLKEAVAVTVEBANQOE-----VAS 207
QY 211 LQQTNGAPA-----SPSELPGTGTAS-----SKPLPLPFS-----PI 243
DB 208 PEEDAEPCIDPSGQSGIOAENMVSGLKFPQDRISKDSQPLRLDDGVDEECRA 267
QY 244 KVVQTSKTLASSSSSSSSSGAERPKQAFILYHPLGSHPPSAGV---GRSHKPPAP 300
DB 268 Q-AQMPPELOSNSTQDDPPDS-----GASFSSSARKTORPHYLAKKLHSP 312
QY 301 SPALPGSTQOLIASPHLAFSTTGLLAAQSLGARGLENTASGFLKPKNGSELSTGEV 360
DB 313 SSGCP-----PRAKTEP-----GAQSGPPTLSRSPGPA----- 343
QY 361 MGPLEPRGGRHKRCFKAKVGSDSALQIHLNHSHTGERPKYKCNVCGNRFTTRGLKVFHR 420
DB 344 -GSSPRQGRYRCGECGKAFQDLCHLKNHAFVHTGKRPCTECGKSYSESEEFKAMHLG 402
QY 421 HREKYPHVQKNPVP-----EHLDVYITSSGLRPGMVPREKAEELATPG 467
DB 403 HR-----GVRRPCCPOCDKAYGTQRDIKEH--QVHSGARPPACD---OCGKAFAFRPS 451
QY 468 GGVERRPLVASTALSTESTLTLSTAGTAPGLPAFNKEVLMKAVEPKNADEMTTP 527
DB 452 LRHRR-----TIQVAPARA-----PCRPVCGRPL 477
QY 528 GSGSASISGVAESSTATLMQLSKMTSLPSWALLTNHK-STGSPFL--PLCARAL----- 580
DB 478 ANQGS-----LNHMRMLHNGEKPFLCHGCGAFRORG 509
QY 581 -----GASPS-----TSKIQQLVEKIDROGAVVATSAASGAPTTAPAPSSASS 626
DB 510 NLRGHLRLHTGERYRCRPHCADAPQPLRL--RRLHLSHTGE----- 550
QY 627 GPNQVYCLRLVLSCPRALRLHYGQHGERPFKCYGCAAFSTGNTLRHFGVHKAAPAR 686
DB 551 --HLCPVCGKALRPHRLRAHERLHSGERFPCCGCGRAVTLATKLRHLKSHLEDKPYR 608
QY 687 AONSCPICQKKTNAVNTLQOHVHMLGQIPINGGALPEGGGAQENGSEOSTVSGAGSF 746
DB 609 ----CPYCGMGY----TLPSGLRRHQLSHRPEAPCSPSPVSAASE-----PTVVLQAE 655

QY 747 PQ-----QSOQSPSEELSEEEDEEDVDDEDSLA 782
DB 656 POLDTHREBEVSPARDVEVTISEQKCFVYPEEDPA 695
RESULT 31
ZFP3_MOUSE
ID ZFP3_MOUSE STANDARD; PRT; 645 AA.
AC Q6116.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Zinc finger protein 93 (Zfp-93).
GN ZFP93.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CD-1; TISSUE-Testis;
RX MEDLINE=96207307; Pubmed=8617494;
RA Shannon M., Ashworth L.K., Mucenski M.L., Lamerdin J.E., Branscomb E.,
RT Stubs L.;
RT *Comparative analysis of a conserved zinc finger gene cluster on human
chromosome 19q and mouse chromosome 7.";
RL Genomics 33:112-120(1996).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U46186; AAB03529.1; -.
DR HSSP: P25490; 12NM.
DR MGI: 107611; Zfp93.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 13.
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PR00048; ZINCFINGER.
DR ProDom: PD000003; Znf_C2H2; 11.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; Znf_C2H2; 13.
DR PROSITE: PS50805; KRAB; 1.
DR PROSITE: PS00028; ZINCFINGER_C2H2_1; 13.
DR PROSITE: PS50157; ZINCFINGER_C2H2_2; 13.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 8 86 KRAB.
FT DOMAIN 285 643 ZINC FINGERS.
FT ZN_FING 285 307 C2H2-TYPE.
FT ZN_FING 313 335 C2H2-TYPE.
FT ZN_FING 341 363 C2H2-TYPE.
FT ZN_FING 369 391 C2H2-TYPE.
FT ZN_FING 397 419 C2H2-TYPE.
FT ZN_FING 425 447 C2H2-TYPE.
FT ZN_FING 453 475 C2H2-TYPE.
FT ZN_FING 481 503 C2H2-TYPE.
FT ZN_FING 509 531 C2H2-TYPE.
FT ZN_FING 537 559 C2H2-TYPE.
FT ZN_FING 565 587 C2H2-TYPE.
FT ZN_FING 593 615 C2H2-TYPE.
FT ZN_FING 621 643 C2H2-TYPE.

30	SEQUENCE	645 AA; 73012 MW; 7459CD140F5AF469 CRC64;
	Query Match	5.1%; Score 269; DB 1; Length 645;
	Best Local Similarity	20.1%; Pred. No. 2,5e-05;
	Matches 139; Conservative	60; Mismatches 214; Indels 280; Gaps 23;
1	294 HKPTAPSPALPGSTQDLASPHLAFPSYTGLLAAQCL-----GAANGLEA-TPSPGL	345
2	197 HKRDVPS---SGDCDQVI-----FPMT--LTIQHCYREQKAYOCSSGQEVFSDSPSL	245
3	346 LKPKNGSGSLSGEVWG-----PLEPRG-GRNCRCAKAFGSD	363
4	246 -----ELHQDTLLGKSPVHSTHKDTRHSPSPVRIQPSYHGRKRYWCHEGKGFRQS	297
5	384 SALQTLHRSHTGEPYKVCNVCNMFTRGNLKVHFRHREKRYPHYOMNPHVENDYVI	443
6	298 SALQTHQKRVHTGEPYKVCNVCNMFTRGNLKVHFRHREKRYPHYOMNPHVENDYVI	443
7	444 TSSGLPYGMSVPEKAEEMEAATPGGVERKPLVASTALSTATESITLLSTAGTAPAGL	503
8	342 -----KCEVCGKG-----	349
9	504 PAFNKFVLMKAVPRKKKADNTPFGSGSALSVAESSTATLMQSLKMTSLPSMALLTN	563
10	350 --FTQWHLQAHERRHITGEPYKCGDCGKRFS-----CSSMLHTH	387
11	564 HFKSTGSPFLPLCARALGASPSSETSRLQQLVEKIDROGAVAVTSAAGAPTTAPAPSS	623
12	388 QRVHTEEKYE-C-----NECGKRFSLSGNLDIQRV-----H	419
13	624 ASSGNQCVICLRLVLSCPRLRLHYGHHGERPRCKVCGRAFSRGNLRHAFVHGKASP	663
14	420 TGEPRYKCECGKSGFSSASFSQSHQRVHTGEPYKVCNVCNMFTR-----SSHFLDQRIH	475
15	684 AARQNSPCICOKKFTNNAVTLQOHVRLHIGQIPINGTALPRGGGAQOENGSGQSVVSGA	743
16	476 TGEKRYRCVCGKRFWSLSLHSHQSVHTGK-----PKCECGKGFHSAS-----	522
17	744 GSPFOQSOQSPSEELSEEEDEEEDVDTEDSLGRSGESGGEKALSVRGDSEEA	803
18	523 -----SLQAHNSVHTGEPYKFCNVQCKQF	546
19	804 SGAESEVGTVAATAATAGKEMDSNEKTTQOSSLPPRRPDSLDQROPMEGSSGVLGKEE	863
20	547 S-----KTS-----NIOAHQVRHNGE-----	562
21	864 GKKPRSSSPASALTPDEGATSVTLVEELSLQEAARKPESSSSKACVCGQAPPSQAA	923
22	563 --KPYKCDTCGKAFS-----OKSSLQVHORIHITGEPYK--CECGKREFRWSVG	607
23	924 LEHQKTHPKEGPLFTCVFCRQGFLEKATLKKH	956
24	608 LSSHQRVHTGEPYKFCNVQCKGSGFSSQASYFMMH	639
RESULT 32		
25	2N45_HUMAN	STANDARD; PRT; 682 AA.
26	002386; P78472;	
27	01-OCT-1993 (Rel. 27, Created)	
28	01-NOV-1997 (Rel. 35, Last sequence update)	
29	16-OCT-2001 (Rel. 40, Last annotation update)	
30	Zinc finger protein 45 (BRC1744).	
31	ZNF45.	
32	Homo sapiens (Human).	
33	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
34	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
35	NCBI_TaxID=9606;	
36	SEQUENCE FROM N.A.	
37	MEDLINE=97220056; PubMed=9067431;	
38	Constantinou-Deltas C.D., Bashirides E., Patsalis P.C.,	
39	Hadjimarcou M., Kriosei P.M., Ioannou P.A., Roses A.D., Lee J.E.;	

RT	*Complete coding sequence, exon/intron arrangement and chromosome
RL	location of ZNF45, a KRAB-domain-containing gene";
RM	Cytogenet. Cell Genet. 75:230-233(1996).
RN	[2]
RP	SEQUENCE OF 1-400 FROM N.A.
RX	MEDLINE=92217982; PubMed=1559709;
RY	Constantinou-Deltas C.D., Gilbert J., Bartlett R.J., Herbstreith M.,
RA	Roses A.D., Lee J.E.;
RB	"The identification and characterization of KRAB-domain-containing
RC	finger proteins.";
RD	Genomics 12:581-589(1992).
RE	-I- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
RF	-I- SUBCELLULAR LOCATION: Nuclear.
RG	-I- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
RH	FINGER PROTEINS.
RI	-I- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
RJ	-----
RK	This SWISS-Prot entry is copyright. It is produced through a collaboration
RL	between the Swiss Institute of Bioinformatics and the EMBL outstation -
RM	the European Bioinformatics Institute. There are no restrictions on its
RN	use by non-profit institutions as long as its content is in no way
RO	modified and this statement is not removed. Usage by and for commercial
RP	entities requires a license agreement (See http://www.isb-sib.ch/announce/
RQ	or send an email to license@isb-sib.ch).
RR	-----
RS	EMBL; L75847; AAB05653.1; -
RT	EMBL; M67509; AAA36133.1; -
RU	EMBL; M67509; AAA36134.1; -
RV	HSSP; P08046; 1A1I.
RW	TRANSFAC; T04988; -
RX	Genew; HGNC:13111; ZNF45.
RY	MIM; 601781; -
RZ	InterPro; IPR001909; KRAB.
SA	InterPro; IPR000822; Znf_C2H2.
SB	Pfam; PF00096; zf_C2H2; 15.
SC	Pfam; PF01352; KRAB; 1.
SD	PRINTS; PR00048; ZINCFINGER.
SE	ProDom; PD000003; Znf_C2H2; 13.
SF	SMART; SM00349; KRAB; 1.
SG	SMART; SM00355; Znf_C2H2; 15.
SH	PROSITE; PS50805; KRAB; 1.
SI	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.
SJ	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 18.
SJ	transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
SK	Nuclear protein; Repeat; Polymorphism.
SL	DOMAIN
SM	8
SN	78
SO	192
SN	662
SN	214
SN	298
SN	326
SN	332
SN	354
SN	382
SN	388
SN	410
SN	438
SN	446
SN	472
SN	494
SN	522
SN	528
SN	550
SN	578
SN	584
SN	606
SN	612
SN	634
SN	640
SN	662
SN	187
SN	255
SN	255
SN	299
SN	299
SN	303
SN	303
SN	504
SN	504
SN	682 AA; 78242 MW; 9C8E536EB00FCBB CRC64;

Query Match 5.1%; Score 269; DB 1; Length 682;
Best Local Similarity 22.9%; Pred. No. 2,6e-05;
Matches 148; Conservative 50; Mismatches 246; Indels 202; Gaps 25;

365 EKPGRHRCRCFAKVEGSDSALQIHLRSHTEGPRPKCNVCGNFRFTTGNLKVH-FHHR 422
161 EKP---YKGEHCYKSFSSWSSHLQINQRAHAGEKPYKCEKCDNAFRFSSILQAHQVRHSA 217
423 EKYPHQNMNHPREHLDYVTS--SGLPYGMSP---PEKAEELNATPGGVE-RKPL 475
218 KSYTN-----DASYNSEQSQRSLPHQVRPTGENPYKYECCGRNVKSSHCAPL 267
476 VASTATLSATESITLLSTAGTATAPGLPAFNKFYLMKAVEPKNKADENTPPSEGSAIS 535
268 IVHTG-----EKRYKEE-----C 281
536 GVAESSTATLMQSLKMLTSLP-----SW-ALLTNHRK-STGSFPLPLCARALGA 582
282 GVGFSORSYLDVHLKLVHTGKKPKYKCECGKSGKSMRSRLQAHQIHTEGPKYKCECGKFCR 339
583 SPSETKLQ-----QLVKKIDRQ-VAVNTS---AASGAPTTAPAPSSSASSG--- 627
340 SFYSYSHLNIHCRITHTGKPYKCECGKSGKFSVSHLQAHQISHTEGPKYKCECGKFCR 399
628 -----PQCVCICLRVLSCPRALRLHYGQHGGERPFKCYGGRAFSTGNL 672
400 ASNLDRHQRHTGKPRYQCQDCKGKGFSSSDNINFRVHTGKPYKCECGKSGKSOANL 459
673 RAHFVGHKSPARAQNSCPICQKFKETNAVTLQOHVHMLGGQIPNGTALPEGGAQOE 732
460 LAHQRHHTGKPRYK---CGTCGKGFSSSDNINFRVHTG---EKRYKCE 504
733 NGEOSTVSAGSPQQOOSQPSPEBELSEEEDEEEDVTDSDSLAGRGESEGEK 792
505 CGK-----AFSQSSSLQVHORVHTGKPYQCAE-----CGKGSVGSOL 543
793 AISRGDSEBASAEEVGTVAATAATAGKEMDSNEKTOOSSLP PPPPSLDQPOPEQ 852
544 QAHQRCHTGKPRYKCECGKFCR----- 568
853 GSSGVLC--GKEGCGKBERSSSPASALTPREGATSVTLVEELSLQEMRKPEGSSSKKA 910
569 --SNFLAHRGVHTGKPYRC-----DVCGKRFORSRYLQAHQVRHTGKPRYK-- 613
911 CEVCGAQPQOAALEEHOKTHPKESRPLFTVCRCRGSFLERATLKH 956
614 CEBCGKVFSSWSSYLDQAHQVRHTGKPRYKCECGKSGKFSWSSSLI 658

RESULT 33
2184_HUMAN STANDARD: PRT: 751 AA.
AC Q99676; Q60792;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN ZNF184.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RX MIM:602277;
RA Jazwinska E.C.;

*Identification of a novel Krueppel-related zinc finger gene (ZNF184)
RT mapping to 6p21.3".
RL Genomics 40:486-489(1997).
CC - FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC - SUBCELLULAR LOCATION: Nuclear (Probable).
CC - TISSUE SPECIFICITY: PREDOMINANT EXPRESSION IN TESTIS.
CC - SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
CC - SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; AL021918; CAA17278.1; -.
DR EMBL; U66561; AAC51180.1; -.
DR HSSP; P08046; 1A1H.
DR Genew; HGNC:12975; ZNF184.
DR MIM; 602277; -.
DR InterPro; IPR01909; KRAB.
DR InterPro; IPR00822; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 19.
DR Pfam; PF01352; KRAB; 1.
DR PRINTS; PRO0048; ZNCFINGER.
DR PRODom; PD000003; Znf_C2H2; 18.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 19.
DR PROSITE; PSS0805; KRAB; 1.
DR PROSITE; PSS0028; ZINC_FINGER_C2H2_1; 19.
DR PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 19.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 28 99
FT FT 222 748 KRAB.
FT ZN_FING 222 244 ZINC_FINGERS.
FT ZN_FING 222 244 C2H2-TYPE.
FT ZN_FING 250 272 C2H2-TYPE.
FT ZN_FING 278 300 C2H2-TYPE.
FT ZN_FING 306 328 C2H2-TYPE.
FT ZN_FING 334 356 C2H2-TYPE.
FT ZN_FING 362 384 C2H2-TYPE.
FT ZN_FING 390 412 C2H2-TYPE.
FT ZN_FING 418 440 C2H2-TYPE.
FT ZN_FING 446 468 C2H2-TYPE.
FT ZN_FING 474 496 C2H2-TYPE.
FT ZN_FING 502 524 C2H2-TYPE.
FT ZN_FING 530 552 C2H2-TYPE.
FT ZN_FING 558 580 C2H2-TYPE.
FT ZN_FING 586 608 C2H2-TYPE.
FT ZN_FING 614 636 C2H2-TYPE.
FT ZN_FING 642 664 C2H2-TYPE.
FT ZN_FING 670 692 C2H2-TYPE.
FT ZN_FING 698 720 C2H2-TYPE.
FT ZN_FING 726 748 C2H2-TYPE.
FT CONFLICT 27 27 A -> S (IN REF. 2).
FT CONFLICT 295 295 Q -> R (IN REF. 2).
FT CONFLICT 359 359 E -> G (IN REF. 2).
FT CONFLICT 379 381 OHQ -> PHP (IN REF. 2).
FT CONFLICT 498 498 R -> G (IN REF. 2).
FT CONFLICT 526 526 Q -> G (IN REF. 2).
FT CONFLICT 533 533 K -> E (IN REF. 2).
FT CONFLICT 547 549 KHE -> OHQ (IN REF. 2).
FT CONFLICT 746 746 R -> K (IN REF. 2).
SQ SEQUENCE 751 AA; 86174 MW; 249DFB302C6BC97D CRC64;

Query Match 5.1%; Score 268; DB 1; Length 751;
Best Local Similarity 21.2%; Pred. No. 3,2e-05;
Matches 136; Conservative 43; Mismatches 220; Indels 242; Gaps 21;
365 EKPGRHRCRCFAKVEGSDSALQIHLRSHTEGPRPKCNVCGNFRFTTGNLKVH-FHHR 424

```
Db 303 EKP---YKCECCGAFSQRTHLYVQHRIHTGKPYKNECGKAFFSORG-----HFMHQK- 354
2Y YPVQWMPHPVPEHLIDYVITSSGLPYGMSVPEPKAEFEAATPGGVRKRLVASTTALSA 484
Db 355 -HTGKPEFCDE-CDKTFYTS-----THLTQ 379
2Y 485 TESL-----TLSTAGATATAPGLPAFNKFLMKAVEPKRKADENTPPGSEGAISGVAE 539
Db 380 HOKIHGKETYKKECGKA-FNGPSTFIRHMIHTGKPEYEC-----CGNAF 427
2Y 540 SSTATLMQLSKLM-----SLPVALLNHFK-STGSPFLPL--CARAL----- 580
Db 428 SOHSNLTQHKHTHGKRYDCAECGKSFYSWSLSLAQHLKHTGKPYKNECGKAFFSYCS 487
2Y 581 -----GASPETSILQOLVEKIDROGAVAVISAAGAPTTSAPAP 620
Db 488 SLTQHRHRIHTREKPEFSECGKATSYLSNLNOHQHTQOKAIECKECCGKAFTRSSSLAK 547
2Y 621 SSSASSG--PNOVCILRVLSCPRLRLHYGOHGERPRPKCVGAFSTRGNLRAHFVG 678
Db 548 HERIHGKERYQCHEGCKTFSGYSSLIOHRKIHTGERPYKNECGNAF-----NQHILTQ 603
2Y 679 HKASPARARQNSCPIQCKKFTNNAVTLQOHVHMLGQITNGGIALPEGGGAOENGEBOG 738
Db 604 HKRIHGAKPYECACGKAFFHSCSLAQHKTH----- 636
2Y 739 TVSGAGSFPOQSQPPEELSEEEDEEEDVDDEDLAGRSESGEKALISVRG 798
Db 637 -----TEKKP----- 641
2Y 799 DSEBAGAEVEVGTAAATAAGKEMDSNEKTQGSLLPPPPPSLDQPMQSGSSVL 858
Db 642 -----YQCNKCEKTFSGS-----HLTQHRIHGE---- 667
2Y 859 GGEKEGKPRSSSPASALTPGEFATSVTLVEELSLQEMRKKEPSSSKAKAEVCCQAF 918
Db 668 -----KPYKNECDKAFS-----RSTHLEHQNHTTGKPYN--CNECKTF 707
2Y 919 PQGALPEHQHTPKEGPLFTVCVCRQGFLEKATLKKHML 959
Db 708 SQSTYLIQHRIHSGKP-FGCNDCGKSFRRSALNKHQL 747

RESULT 34
ID ZOT1_XENLA STANDARD; PRT; 898 AA.
AC P18751;
2T 01-NOV-1990 (Rel. 16, Last sequence update)
2T 01-NOV-1990 (Rel. 16, Last sequence update)
2T 01-OCT-1994 (Rel. 30, Last annotation update)
2E Oocyte zinc finger protein XLOC07.1 (Fragment).
2C Xenopus laevis (African clawed frog).
2C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
2C Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
2C Xenopodinae; Xenopus.
2C NCBI_TaxID=355;
2X [1]
2N SEQUENCE OF 1-311 FROM N.A.
2P MEDLINE=9040698; PubMed=2503827;
2X MEDLINE=89345612; PubMed=2503827;
2A Knoechel W., Poeling A., Koester M., el Baradi T., Nietfeld W.,
2A Bouwmeester T., Pieler T.,
2E "Evolutionarily conserved modules associated with zinc fingers in
2T Xenopus laevis.";
2T Proc. Natl. Acad. Sci. U.S.A. 86:6097-6100(1989).
2N [2]
2P SEQUENCE OF 284-898 FROM N.A.
2X MEDLINE=90040698; PubMed=2509712;
2A Nietfeld W., el Baradi T., Mentzel H., Pieler T., Koester M.,
2A Poeling A., Knoechel W.;
2T "Second-order repeats in Xenopus laevis finger proteins.";
2T J. Mol. Biol. 208:639-659(1989).
2C -----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M25866; AAA50013.1; -.
DR PIR; A33282; A33282.
DR PIR; S06546; S06546.
DR HSSP; P08047; ISP2.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 21.
DR SMART; SM00355; znf_C2H2; 21.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 21.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 21.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT ZN_FING 289 311
FT ZN_FING 342 364 C2H2-TYPE.
FT ZN_FING 370 392 C2H2-TYPE.
FT ZN_FING 398 420 C2H2-TYPE.
FT ZN_FING 425 448 C2H2-TYPE.
FT ZN_FING 426 448 C2H2-TYPE.
FT ZN_FING 454 476 C2H2-TYPE.
FT ZN_FING 481 504 C2H2-TYPE.
FT ZN_FING 482 504 C2H2-TYPE.
FT ZN_FING 510 532 C2H2-TYPE.
FT ZN_FING 537 561 C2H2-TYPE.
FT ZN_FING 538 561 C2H2-TYPE.
FT ZN_FING 567 589 C2H2-TYPE.
FT ZN_FING 595 617 C2H2-TYPE.
FT ZN_FING 623 645 C2H2-TYPE.
FT ZN_FING 651 673 C2H2-TYPE.
FT ZN_FING 679 701 C2H2-TYPE.
FT ZN_FING 707 729 C2H2-TYPE.
FT ZN_FING 735 758 C2H2-TYPE.
FT ZN_FING 764 786 C2H2-TYPE.
FT ZN_FING 792 814 C2H2-TYPE.
FT ZN_FING 820 842 C2H2-TYPE.
FT ZN_FING 848 870 C2H2-TYPE.
FT ZN_FING 876 898 C2H2-TYPE.
FT NON_TER 898 898 C2H2-TYPE.
SQ SEQUENCE 898 AA; 101167 MW; 795D8065696B06 CRC64;

Query Match 5.1k; Score 268; DB 1; Length 898;
Best Local Similarity 19.3k; Pred. No. 3.8e-05;
Matches 201; Conservative 91; Mismatches 371; Indels 378; Gaps 40;

QY 17 EPALGDASEEDHPQVCAKCAQFTDPTFLAHQNA---CSTDPPVMYIIGQENPNNS 73
Db 126 EPQQL-----RPQC-----EYKDESNVAHMEATLCNSD-----GNFINENP 165
QY 74 --SASSPRREGHNPNQVMDTEHNSPPRSGSSVPPDPPTWGPERRGESSGHFLVAATGTA 131
Db 166 EISPGEQPPRANGIKKEATSEEGNSQCSINPLLEEIOGTD----- 207
QY 132 AGGGGGLILASPKLGATPLPESTPAPPPPPPPPGVSGHLNIPL-----ILLEE 183
Db 208 -----TPTP-----IMGSLNLSLDNYISDEKKE 233
QY 184 LRYLQQRQIHOMQNTQICQVLLLSIGQTVGAPASPELPGTASTKPLPLPSPT 243
Db 234 ETSQGRNQSDCSIT-----PLTEQIQGTDPTPIMGCSLKDNKYDGNPHMSPK 282
QY 244 KPYQT-----SKTLASSSSSSSSSGAETPQAQAFPHLY----HPL 279
Db 283 NTLRRRYSCNECHEYLLHKRDFGKHQWTHKREKSFSCSECG-----KCPNLQCPDRHQPT 337
QY 280 GSOHPSPSAGVGGRSHKPTPAVSPALPGSTDLIASPHLAFPST-----TGLLAQCL 331
Db 338 GK-KPSCSKGKC-----FAFLSDLIVHRIHTGERPPSCS 373
```

332 GAARGLLEATASPELL-----KPKNGSGELSYGEV-----MCPLEKPGGR 370
 Db ECGKGF-----TRPALLIHHHTHTGKPFSCSECGKFSKSSLVHQRTHIG--EKP--- 425
 371 HKCRCAKAVGSDSALQILRSHTGERPKYCNVCGRNFTTGNLKVHFRHREKRYPHVM 430
 426 FCCSECDKCFASSELINIHQRTHTGKPFSCSECGKCFTHNS-----HFAHHQ----- 473
 431 NHPVPEHLDYITSSGLPRGMSVPRPEKAEDEAATPGGVVEKPLVASTALSTESLT 490
 474 -----MHTG-----BKPFCCKCGKCFASSSDL 497
 491 ILSAGTATAPGLPAFKFVLMKAVEBKNADENTPGSE--GSAISGVESSTATLMQ-- 547
 498 -----TFHR-----RTHREKTFSCSECGKCFSNHSLAHQMHQHTG 534
 548 -----LSKMTSLPSWALTLNHFSTGSPPLPCARALGASPESTKL--QOLVERIDRQ 600
 535 EKPFCCSECGKCFSSSGLTAHQRTHKVKPFSCSACGKCFSNRSHLIRHQMHGK- 593
 601 GAVAVTSAASGATTSAPARSSASSGPNOCVLCRLVLSCPALRLHYGONGEPRPKCK 660
 594 -----PFCSECKKCFSPNSNLARHQMTHTGKPFSCS 626
 661 VCGRASTGNLRAHFVGHKASPARAQNCPICQKFTNATVLOQHVRLHGLQIPNGG 720
 627 ECGKCFASSSDLTFHHHTH---TGKPFSCSECGKCFYKSKSSLVHQRHTH--GEKPFSC 681
 721 TALPEGGAQAENGSEGSTVSGAGSPFOQSQSPSEBELSEEEDEDEEEDVTDSDS 780
 682 SKCDKCFASSELINIHQRTHTGKAF-----SCSECGKCFTHRSQ 721
 781 LAGRGSESGEKALISVGDSEAGAEVEVTAATAATGKEMDSNEKTKQSSLPFRPP 840
 722 LSRHQMHGTEKPIISC--PECEECF-----VSSQLTAHQQAANRNVK 762
 841 PSLDPOQPMQEGSSVGLGKEEGK--PERSSPASALTPEGEATSVTLVEELSLQEAM 898
 763 PRSC-----LECGKCFSNRNSNFRHQMHG-----KPFSCSECKR 798
 899 RKEPGESSSRK-----ACEVCGQAPRPSQALAEHQKTHPEKGLFTVCFCROGF 947
 799 KGRSNOSSLARHQMTHTGKPFSCSECGKCFRNSQSLARHQMHGKPF--FSCSECAKGF 857
 948 LERATLKKHMLLAHHVOFPA 968
 858 SNOGLARHQ-MTHTEKPPA 877

RESULT 35
 2192_HUMAN STANDARD; PRT; 578 AA.
 AC Q15776; Q9H4T1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Zinc finger protein 192 (ZNF192).
 GN ZNF192.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT LEU-163.
 RC TISSUE=Ovary;
 MEDLINE=97386587; PubMed=9244436;
 RA Lee P.L., Gelbart T., West C., Adams M., Blackstone R., Beutler E.;
 RT "Three genes encoding zinc finger proteins on human chromosome 6p21.3:
 RT members of a new subclass of the Kruppel gene family containing the
 RT conserved SCAN box domain";
 RL Genomics 43:191-201(1997).
 RM [2]
 RP SEQUENCE FROM N.A.

RA Williams S.;
 RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP PARTIAL SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=96230927; PubMed=8673473;
 RA Beutler E., Gelbart T., West C., Kuhl W., Lee P.;
 RT "A strategy for cloning the hereditary hemochromatosis gene";
 RL Blood Cells Mol. Dis. 21:206-216(1995).
 CC - FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
 CC - SUBCELLULAR LOCATION: Nuclear (potential).
 CC - SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC - SIMILARITY: CONTAINS 1 SCAN BOX.
 CC - SIMILARITY: CONTAINS 1 SCAN BOX.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U57796; AAB02260.1; -;
 CC EMBL: AL358933; CAC15901.1; -;
 CC EMBL: U88080; AAC51656.1; -;
 CC EMBL: U88079; AAC51656.1; JOINED.
 CC HSSP: P07248; 1PAA.
 CC Genew: HGNC:12983; ZNF192.
 CC MIM: 602240; -;
 CC InterPro: IPR001909; KRAB.
 CC InterPro: IPR003309; Treg_SCAN.
 CC InterPro: IPR000822; Znf_C2H2.
 CC Pfam: PF00096; Zf-C2H2; 9.
 CC Pfam: PF01352; KRAB; 1.
 CC Pfam: PF02023; SCAN; 1.
 CC PRINTS: PR00048; ZINCFINGER.
 CC ProDom: PD000003; Znf_C2H2; 8.
 CC SMART: SM00349; KRAB; 1.
 CC SMART: SM00431; LER; 1.
 CC SMART: SM00355; Znf_C2H2; 9.
 CC PROSITE: PS50805; KRAB; 1.
 CC PROSITE: PS50804; SCAN_BOX; 1.
 CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.
 CC PROSITE: PS00157; ZINC_FINGER_C2H2_2; 9.
 CC Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 CC Nuclear protein; Repeat; Polymorphism.
 CC DOMAIN 51 133
 CC DOMAIN 220 316
 CC FT DOMAIN 322 568
 CC FT ZN_FING 322 344
 CC FT ZN_FING 350 372
 CC FT ZN_FING 378 400
 CC FT ZN_FING 406 428
 CC FT ZN_FING 434 456
 CC FT ZN_FING 462 484
 CC FT ZN_FING 490 512
 CC FT ZN_FING 518 540
 CC FT ZN_FING 546 568
 CC FT VARIANT 163 163
 CC FT CONFLICT 34 34
 CC FT CONFLICT 113 113
 CC FT SEQUENCE 578 AA; 65757 MW; 5914F5439F4B8AB8 CRC64;
 SQ
 Query Match 5.1%; Score 267; DB 1; Length 578;
 Best Local Similarity 22.7%; Pred. No. 2.7e-05;
 Matches 137; Conservative 61; Mismatches 207; Indels 198; Gaps 23;
 OY 177 IPILELRVLIQQRIOHQMOTEOIC-----ROYLLLS-----LQTVAP 218
 Db 100 LTIPEELQTLVKD--HOLENCEEVVTLLEDLERQIDILGRPVASRVHGRVLEEVVHS 157

y		219	ASPSLPLPG---CTAASKPRLDLPSPKRVQTSKITLAASSSSSSS-----SSGAE	266
y	:	:	: : : : : : : : : : : : : : : : :	:
b	158	AAPPPPTOLIOSEATVQHKSVPVOESOERAMSTQSOPRSCSGSDGDEMTATLLTAGPO	217	
T-	267	T-----PKQA-FEHNLHPICGSQPFFSAAGVGGRSHKPRTAPSPALPG	306	
y	218	TLEKIEDMAVSILIREWMLIDPSQKDLCRDNRPEHFRNMFSLGGFETRSNNRELA-----	270	
b	307	STDQLIAS---PHLAFPTTGLLAACLG--AAKGLETASPGLKKPKNGSELGTGEWG	362	
y	:	:	:	:
b	271	-SKOYISTGIQIH-----GETPAKNCGDVIRGLEHEEARDDL-----G	307	
y	363	PKEKPGG-----RHKCRCFAKYFGSDALOHLRSHTEGERPYCNVCGNRPTRTNLKV	416	
Ib	308	RUERRGNTQTGERHNHCDCCKGFAPSLSGLVRHMRIHTGERPYCNCVKGAFTYSALLS	367	
y	417	H-FHRHKRYPHYVMNPDPVEPHLDIVITSSGLIPYGMSVPEKAEEEAAPPDGVERRK	474	
Ib	368	HODINNKVKRY-HCK-----ECGKFASFONT-----GHILHOR	398	
y	475	LVASTVALSATSETLTLLISASTATAAGLPANFNFLMKAVEPNKKADENTPPGESGAI	534	
Ib	399	IHTGEKPYCNCVKGFASOSAG-----LIHQRIHSGERPYECNE-----	438	
y	535	SGVASSTATLMQISKIMTSLPFWALLTHFKSYSGSFPLLCAHALGASPFSTSKIQQLY	594	
Ib	439	CCKAKFSHSLIGHQRHT-----GEPYECDCEGKTFE-----	471	
y	595	EK----IDQGAVAYVNASASGAPTTSPAARSSASSGNPCVICLRVLSCPRAFLNHQY	650	
Ib	472	RSSHILIGHQ-----RSHTGEKPYKKNCCGRAFSQSGSLIEHQRI	511	
y	651	HGGERPKFCCKCVKGVSTRGNLRAFHVGHKASPARARONSCPICOKKTFNAVTLOOHWM	710	
Ib	512	HGERGYKCKECKGKANNGNTGLIQLHKLHIH-----TGEMKYQCNEGKAPIQRSILLRHORI	567	
Oy	711	HIG 713 .	.	
Ib	568	HSG 570		
<hr/>				
RESULT 36				
ZFP95_HUMAN				
ID	ZFP95_HUMAN	STANDARD;	PRT; 839 AA.	
AC	Q9ZEL8;			
Dt	16-OCT-2001 (Rel. 40, Created)			
Dt	15-OCT-2001 (Rel. 40, Last sequence update)			
Dt	15-JUN-2002 (Rel. 41, last annotation update)			
De	zinc finger protein zfp-95.			
Gn	ZFP95 OR KIAA1015.			
Os	Homo sapiens (Human).			
Oc	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Ox	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
Nr	NCBI_TaxId=9606;			
Rf	[1]			
Rp	SEQUENCE FROM N.A.			
Rp	MEDLINE=20054365; PubMed=10585779;			
Ra	Dreyer S.D., Zheng Q., Zabel B., Winterpacht A., Lee B.;			
Rt	"Isolation, characterization, and mapping of a zinc finger gene,			
Rt	ZFP95, containing both a SCAN box and an alternatively spliced KRAB A			
Rl	domain.";			
Rl	Genomics 62:119-122(1999). [2]			
Rn	SEQUENCE FROM N.A.			
Rp	TISSUE=Brain;			
Rc	MEDLINE=99246063; PubMed=10231032;			
Ra	Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,			
Rt	Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;			
Rt	"Prediction of the coding sequences of unidentified human genes.			
Xiii.	The complete sequences of 100 new cDNA clones from brain which			
Data	code for large proteins in vitro".			
Res.	6:63-70(1999).			

CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SCAN BOX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF170025; AAF24219.1; -;
DR EMBL: AB023232; BAA76659.1; -;
DR HSSP: P07248; ZADR.
DR TRANSFAC: T04978; -;
DR Genew: HGNC:12867; ZFP95.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR003309; Treg_SCAN.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PR00096; zf-C2H2; 12.
DR Pfam: PF01352; KRAB; 1.
DR Pfam: PF02023; SCAN; 1.
DR ProDom: PD000003; Znf_C2H2; 9.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00431; LER; 1.
DR SMART: SM00355; ZNF_C2H2; 12.
DR PROSITE: PS50805; KRAB; 1.
DR PROSITE: PS50804; SCAN_BOX; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 13.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 51 132 SCAN_BOX.
FT FT 218 291 KRAB.
FT FT 346 823 ZINC_FINGERS.
FT FT ZN_FING 346 368 C2H2-TYPE.
FT FT ZN_FING 374 396 C2H2-TYPE.
FT FT ZN_FING 402 424 C2H2-TYPE.
FT FT ZN_FING 430 452 C2H2-TYPE.
FT FT ZN_FING 549 571 C2H2-TYPE.
FT FT ZN_FING 577 599 C2H2-TYPE.
FT FT ZN_FING 603 627 C2H2-TYPE.
FT FT ZN_FING 633 655 C2H2-TYPE.
FT FT ZN_FING 661 683 C2H2-TYPE.
FT FT ZN_FING 717 739 C2H2-TYPE.
FT FT ZN_FING 745 763 C2H2-TYPE (ATYPICAL).
FT FT ZN_FING 773 795 C2H2-TYPE.
FT FT ZN_FING 801 823 C2H2-TYPE.
SQ SEQUENCE 839 AA; 96903 MW; F2E062CF6610806D CRC64;

Query Match 5.0%; Score 266; DB 1; Length 839;
Best Local Similarity 19.9%; Pred. No. 4.3e-05;
Matches 121; Conservative 76; Mismatches 199; Indels 212; Gaps 22;

QY 371 HRCRFAYKVGSDALDILHRSHTGERPKYKVCNGNRFTRTGLKLVHFNHREKYPH--- 427
DB 346 HRCSDGKFKFLQASNFLOHRRHILGKPFKCEGCKSYQNRVHLLTDHQRVHTGEKPKCQ 405
QY 428 -----VQMNHPVPEHLDVYTTSSGLPRGMSVPRPEKAEAEATPGGVEKRPVASTTAL 482
DB 406 VCGKARVVSLLVQHH---SVHSGERPLGCN-----ECGKNFG--RHSLLIEH----- 448
QY 483 SATESTLLSTASGATATAPGLPAFNFKFLMAVEERKNKADEMTPPSGEGSALGVASST 542
DB 449 -----LKRHRERKQSDSKRSKNT-----KLSVKKKISYESE 481
QY 543 ATLMOISKLMTSLPSWALLTNHFKSTGSPPLPLCARAGASPSPTSKIQDLVE-----K 596
DB 482 AD-MEIS-----GKTQNVASQVDPFGEGCEFGQK 509


```
2Y 965 QPF 967
b 680 RPY 682

RESULT 38
7335_HUMAN
ID 2335_HUMAN STANDARD: PRT; 1342 AA.
Q9H422; Q9H684;
15-JUN-2002 (Rel. 41, Created)
15-JUN-2002 (Rel. 41, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
DE ZINC finger protein 335.
ZNF335.
ON Homo sapiens (Human).
DC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
KA MEDLINE-21636749; PubMed-11780052;
KA DeLounas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
KA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
KA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
KA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
KA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
KA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
KA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
KA Coulson A.G., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
KA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
KA Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
KA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
KA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
KA Kay M.P., Kimberley A.M., King A., Knights A., Laidl G.K., Lawlor S.,
KA Levenshlo M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
KA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McMurray A.A.,
KA Oliver K., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
KA Phillimore B.J.C.T., Pratchalingam S.R., Plumb R.W., Ramsey H.,
KA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
KA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
KA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
KA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
KA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
KA Whiting L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
KA Rogers J.;
JT The DNA sequence and comparative analysis of human chromosome 20.;
NT Nature 414:865-871(2001).
[2]
SEQUENCE OF 455-1342 FROM N.A.
1 KAWABATA A., HIKIJI T., KOBATAKE N., INAGAKI H., IKEMA Y., OKAMOTO S.,
2 OKITANI R., OTA T., SUZUKI Y., ODAYASHI M., NISHI T., SHIBAHARA T.,
3 TANAKA T., NAKAMURA Y., ISOGAI T., SUGANO S.;
4 "NEDO human cDNA sequencing project.";
5 Submitted (JUG-2000) to the EMBL/Genbank/DBJ databases.
6 -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
7 -1- SUBCELLULAR LOCATION: Nuclear (Probable).
8 -----
9 This SWISS-PROT entry is copyright. It is produced through a collaboration
10 between the Swiss Institute of Bioinformatics and the EMBL Outstation
11 at the European Bioinformatics Institute. There are no restrictions on its
12 use by non-profit institutions as long as its content is in no way
13 modified and this statement is not removed. Usage by and for commercial
14 entities requires a license agreement (See http://www.isb-sib.ch/announce/
15 or send an email to license@isb-sib.ch).
16 -----
17 EMBL: AL162458; CAC10457.1; -.
18 EMBL: AK026157; BAB15379.1; ALT_INT.
19 Genew; HGNC:15807; ZNF335.
20 InterPro: IPR000822; Znf_C2H2.
21 Pfam: PF00006; Zf-C2H2.21.
22 ProDom: PD000003; Znf_C2H2; 1.
23 SMART: SM00355; Znf_C2H2; 13.
24
```

```
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 13.
KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;
KW DNA-binding; Repeat.
FT ZN_FING 245 268
FT ZN_FING 465 487 C2H2-TYPE.
FT ZN_FING 495 517 ZN_FING.
FT ZN_FING 523 545 C2H2-TYPE.
FT ZN_FING 562 584 C2H2-TYPE.
FT ZN_FING 590 612 C2H2-TYPE.
FT ZN_FING 621 643 C2H2-TYPE.
FT ZN_FING 649 672 C2H2-TYPE.
FT ZN_FING 678 701 C2H2-TYPE.
FT ZN_FING 1019 1041 C2H2-TYPE.
FT ZN_FING 1047 1069 C2H2-TYPE.
FT ZN_FING 1075 1097 C2H2-TYPE.
FT ZN_FING 1103 1126 C2H2-TYPE.
FT DOMAIN 1178 1350 GLN-RICH.
SQ SEQUENCE 1342 AA; 144892 MW; 6D230DE0B3AB670 CRC64;

Query Match 5.0%; Score 265.5; DB 1; Length 1342;
Best Local Similarity 18.7%; Pred. No. 7.2e-05;
Matches 234; Conservative 99; Mismatches 419; Indels 501; Gaps 45;

QY 16 GEPALGDSAEEDHPQVCAKCAQFTDPTFEFLAHQNACTDPPVYTIIGQENPNSSA 75
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 57 GOSRGRSGSOEE---VSSESSADLPNSYLPDSSVS-HGPVAGVTGCPALVHSSA 111
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 76 SSEPPPEGNINQVMDTEHNSNPDSGVSP--TPPTMCPERKGESSGHFLV--AATGTA 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 112 LPDP-----NMIVSCTGASS-SDLGSAIDKTIESTIGPD-----LIQNCITVTS 154
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 132 AGGGGG---LILASPKGATPLP-----ESTAPP 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 155 AEDGAETTRYLILQPDGAPMTSPSSSTLAHSLAIEALADPTSTSTLEAGGGS 214
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 160 PPPPPPPPGVSGHINPLILELRYLQQRQIHQM-----MTEQICRQVIL 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 215 SPVQLPPASGAEEPDQSLDAMMEVVVVOQFKMCCYRSSTKATLLRHRHRPVAA 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 208 LGSIGQTVG-----APASPELGTG-- 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 275 AAAAGKKGRLLKKNSTSGQEEGPEEEDDDIVDGAIDLDLEDSDIYNRAEDPRGQ 334
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 229 -----TASSTRKPLPLESPIKPVQTSKTLASSSSS-----SSSGAETPKQAFHL 275
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 335 LRLQRPSTPRPR--RPRGRPKLPRLEISDLPDVEGEPVSSGQGPPE----- 385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 276 YHPLGSOHPFSAG-----GVGSHKPTPAAPSPALPGSTQILLASPHLA 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 386 --PQDEAPSSSGPGLVAMGVKSRTPVEAGVSQSDAENAP--CPDEHDTL----- 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 319 FPGTGLAAQCLGARGLEATASPGILKPKNSGELSYGVGMPLFPGGRHKRCFCAK 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 435 -PRRGRGRSRRLKRYKRYKSP-----KPLLRP---FLCRIGGS 472
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 379 VFGSDSAIQIHRSHTGERP--YKQNVCCNRFTTRGLUKYHFRHREKYPHYQMPHPV 436
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 473 RFLSHEDLRFVNSHSEAGDPOLFKLCQSYRSRWSLSIKEMFNH----- 517
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 437 EHLDPVITSSGLPGYGVYPEKAEEEAATPGGVVERPVAASTALSTLSTLSAG 496
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 518 -----VGSAPYKCDECSYT--SVYRKDVIRHAHVH----- 546
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 497 TATAPGLPAFNKFLMKAVEPRKNADENTPPGSGSAISGVAESSTATLMQLSKIMTSLP 556
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 547 -----RDRKKRPDPPTP----- 557
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 557 SWALLTNFKSTGSPPLPCARALGASPSSETSKLQQLVEKIDROGAAVAVYTSAAAGAPTS 616
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 558 -----KLSSPPCVCGRVY--PMOKRLTOHM-----KTHS 585
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 617 APAPSSASSGPN-----QCVICLRVLSCPRALRLHYGGHG 653
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Db	Seq	Accession	Species	Length	Score	E-value	Bit Score	Identities	Positives	Gaps	Conserved Domains	Annotations
Db	586	TEKPMCDKCGASF	KRKYFFKMLLTHIQAVNRRFCEFEFYCEDKALKNLHOLSHVS	645								
QY	654	ERPFCVKVGAFASTR	GNLRANF-VGKH-ASPAAQNSCPTQCKFFNNAVTLQOVRMH	711								
Db	646	DKPFCSCPCPYRTFR	EDFLSHVAVKITGAKPFA-----CEYCHSTHKKKRLRLHVKR	700								
QY	712	LGQIPNGSTAL	PEGGAAQENSGS--TVSAGSPFOQSOQSP-----EEELSE	762								
Db	701	HASFEEMWRHDE	-----EPFSRRRPFSLQIEELKQOHSAPRGPPSPGPEIPP	754								
QY	763	EEEDDEDEEDVTD	DSLAG-----RGSSESGEKAISVAGDSFEASGAEVEVTAANA	816								
Db	755	EATTFQSSAPSL	CSDTLGAGATIIYQOGAEESTAMATGALDILLNNSAQRELGTALQ	814								
QY	817	ATAGKEMD	-----SNEKTTQO-----SLPPPPPD-----	842								
Db	815	VAVVSEDEVAGIAS	RGCGPSPCATQVYVTLHVAERPGGAAAEQGLRPRDLQITLACG	874								
QY	843	-----SLDQPPME	QGS--GVLGKREBCKPERSSSPASALTPEGE---ATSVTV	889								
Db	875	PEGGCGYSVITAP	PMEEGTAPCTPYSEEPAGEAQAQAVVSDTLKEAGTHYIMATDGTQL	934								
QY	890	EELSL	-----QEMRKRP	902								
Db	935	HHIELTAGDIS	ISFPSPDALASGAKWPLLQCGGLRDRDGPSPSPAKTHCGVDSOSSASSPP	994								
QY	903	GRS	-----SSRKAEVCGQAFPPSQALAEHQTHPKKEGFLTCVCF-	943								
Db	995	ATSKALGLAVP	PPSPSAATASKKFSCKICAEAPRGRAEMSHKRAHAGP-AFRCPPDP	105								
QY	944	-----RQGLE	RATLTKHM-----LLAHQVQPPFAPH	970								
Db	1054	FSARQWPEVRAH	AQOHSLSLRPHQCCSQCSFASKNKKDLRRHMLTHRKKEPFACH	1106								

RESULT 39

2337_HUMAN ID 2337_HUMAN STANDARD; PRT; 751 AA.

AC Q9Y3M9; Q9Y3Y5; 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Zinc finger protein 337.

GN ZNF337.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI_Taxid=9606;

CK [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Lloyd G.K., Lawlor S., Leharastaino H.H., Levarsha M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConachie L.J., McIay K., McMuray A.A., Milne S.A., Misty J.D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Pithermore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.M., Rose M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann R.W., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,

```

RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RA "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [2]
RP SEQUENCE OF 256-751 FROM N.A.
RC TISSUE=Brain;
RA Mambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -! SUBCELLULAR LOCATION: Nuclear (Probable).
CC -! SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -! SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL031673; CAB43548.2; -.
DR EMBL; AL049942; CAB43216.1; -.
DR HSSP; P08048; 7ZNF.
DR Genew; HGNC:15809; ZNF337.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 37.
DR Pfam; PF01352; KRAB; 1.
DR Prodom; PD000003; Znf_C2H2; 4.
DR SMART; SM00355; ZNF_C2H2; 19.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 20.
KW Transcription regulation; zinc-finger; Metal-binding; Nuclear protein;
KW DNA-binding; Repeat.
FT DOMAIN 12 83 KRAB.
FT 180 735 ZINC FINGERS.
FT ZN_FING 180 202 C2H2-TYPE (DEGENERATED).
FT ZN_FING 208 230 C2H2-TYPE.
FT ZN_FING 236 258 C2H2-TYPE.
FT ZN_FING 264 286 C2H2-TYPE.
FT ZN_FING 292 314 C2H2-TYPE.
FT ZN_FING 320 342 C2H2-TYPE.
FT ZN_FING 348 370 C2H2-TYPE.
FT ZN_FING 376 398 C2H2-TYPE.
FT ZN_FING 404 426 C2H2-TYPE.
FT ZN_FING 432 454 C2H2-TYPE.
FT ZN_FING 460 482 C2H2-TYPE.
FT ZN_FING 488 510 C2H2-TYPE.
FT ZN_FING 516 538 C2H2-TYPE.
FT ZN_FING 544 566 C2H2-TYPE.
FT ZN_FING 572 594 C2H2-TYPE.
FT ZN_FING 600 622 C2H2-TYPE.
FT ZN_FING 628 650 C2H2-TYPE.
FT ZN_FING 656 679 C2H2-TYPE.
FT ZN_FING 685 707 C2H2-TYPE.
FT ZN_FING 713 735 C2H2-TYPE.
FT CONFLICT 476 476 T -> A (IN REF. 2).
FT CONFLICT 476 476 E -> V (IN REF. 2).
FT CONFLICT 653 653 T -> A (IN REF. 2).
SQ SEQUENCE 751 AA; 86874 MW; 482D9F9A57EC980E CRC64;

Query Match 5.0%; Score 265; DB 1; Length 751;
Best Local Similarity 19.0%; Pred.No. 4,3e-05;
Matches 167; Conservative 92; Mismatches 269; Indels 352; Gaps 28;

```

52 GILHSKPELIRLEQGEVPMGEERRR-----PGPCAGIYAEHVLR---KNLGLAHOR 102
191 QIHOMOMTEOICHOVLLLSLGLQTVGAPASPELBTGTGASSTKPLLPLESPKPYOTSK 250
103 Q-QQLOPDSQFS-----DTAQOKEKSTKPMAFSSPPLAHAVSR 144
251 TLAASSSSSSSSSGAETPKQAFPHLYHPLGSOHPPSAGVGSRHKPTPPASPALPGSTDQ 310
145 RNSVYEIESQOGRENPT-----IDK 167
311 LLASPLAPFTTGILAAOCLGAARGLFATASPGLLKPKNSGELSYGEVMPLEKGGCR 370
168 VL-----KGLE-NSRMGAFKCAERQDGFRRKMVLIHKKASHR 204
371 HK---CRFCAKVGSDSALOIHRSHTGERPKCNCVCGNRFTTRGMLKVFHHRREKYPH 427
205 QKLFTRCQCQGRDESLHLHOMTHHGEKSYVSCVCGRGSFLKALRLRQTH----- 258
428 VOMNPHVPEHLDVITYSSGLPYGMSVPRPEKAEBAATPGGVERKPLVASTALSTATES 487
259 -----SG-----EKPLCKVCGRGYT-- 274
488 LHLSTSGATATAPGLPAFKFKVLMKAVEPKNAKADENTPPGSEBSAISGVASSTATLMQ 547
275 ----- 274
548 LSKLMTSLPSMALLTNHFS--TGSFPLPL--CARALGASPSFETSKLOOLEKIDROGAVA 604
275 -----SKSLITHERTHTGKPYECQCGRRPNDKSSINXKHLK----- 312
605 VTSASGAPTTSPAPASSASG--PNCVITCLRVLSCPRALRLHYGONGGERPFKVC 662
313 -----AHSGEKPFWCKECGRGYTNKSYFVHHRKIRHSGEKPYRCQC 353
663 GRAFSTRGNRAHFVGHKASPARAQNSCPICQKFTNAVITLOOHVBMHGGIIPNGTA 722
354 GGFSN---KSHLTHQTHHSGEPPACROCKOSFSYKSLIRHQTHSGE----- 402
723 LPEGGAOENGSEOSTV-----SGAGSFPOQSOQSPPEELSEEBEEDDEEDV 775
403 -FVPCDCERSFSQKSTLVYHQTHHSGEPPVCRGCGGFIQKSTLVKIQITHSEKPFV 461
776 TDEDSLARGSESGGEKAIVRGDSEASGAEEVGYVAAAATAGKEMDSNEKTTOQSSL 835
462 CKD---CGRGFIQKSTFTLHQTHHSGEPPYCGRCG-----RRFRDKSSYKHL 507
836 PPPPPDSLDQPPMGSSGVJGK---BEGCKPERSSSPASALTTPGEATSVLYEE 891
508 -----RAHLGEKRPFCRDGCR-----GFTLKP 530
892 LSLQEA-----MRKEPGESSSRKA-----CEVGOAFPSQALAEH 927
531 LTIHQTHHSGEPPMCKQCKSFSLKANLRLHQTHHSGEPPNCKCKGCGFILLKSTLLFH 590
928 QKTHPEKGLFTVCFCRQGLERATLKKHMLLAHHVOPE 967
531 QKTHSGEKP-FICSECGQGFIMKSNLVKHQ-LAHSGKOPF 628
RESULT 40
ID DRPL_RAT STANDARD; PRT; 1183 AA.
IC P54258;
JT 01-OCT-1996 (rel. 34, Created)
JT 01-OCT-1996 (rel. 34, Last sequence update)
JT 30-MAY-2000 (rel. 39, Last annotation update)
DE Atrophin-1 (Dentatorubral-pallidolysian atrophy protein).
IN DRPLA.
NS Rattus norvegicus (Rat).
XC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
XC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
XN NCBI_TaxId=10116;
IN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=8541849;
RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA Ashworth R.G., Ross C.A.;
RT "Cloning and expression of the rat atrophin-1 (DRPLA disease gene)
RT homologue."
RN Neurobiol. Dis. 2:129-138(1995).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; PubMed=8541849;
RA Schmitt I., Epplen J.T., Riess O.;
RT "Predominant neuronal expression of the gene responsible for
RT dentatorubral-pallidolysian atrophy (DRPLA) in rat."
RL Hum. Mol. Genet. 4:1619-1624(1995).
CC -I- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
CC -I- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U31777; AA80337.1; -;
DR EMBL; X89453; CAA61623.1; -;
DR InterPro; IPR002951; Atrophin.
DR Pfam; PF03154; Atrophin-1; 2.
DR PRINTS; PR01222; ATROPHIN.
FT DOMAIN 165 171 POLY-PRO.
FT 303 306 POLY-PRO.
FT DOMAIN 377 383 POLY-SER.
FT 387 391 POLY-SER.
FT DOMAIN 440 446 POLY-SER.
FT 447 480 POLY-HIS.
FT DOMAIN 481 489 POLY-GLN.
FT 502 505 POLY-PRO.
FT 562 572 POLY-SER.
FT DOMAIN 702 705 POLY-PRO.
FT 455 455 N -> S (IN REF. 2).
FT CONFLICT 594 594 F -> L (IN REF. 2).
FT CONFLICT 689 689 T -> R (IN REF. 2).
FT CONFLICT 717 717 A -> V (IN REF. 2).
FT CONFLICT 737 737 MISSING (IN REF. 2).
FT 965 965
SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADF9B1F CRC64;
Query Match 5.08; Score 265; DB 1; Length 1183;
Best Local Similarity 21.48; Pred. No. 6.7e-05;
Matches 239; Conservative 94; Mismatches 372; Indels 412; Gaps 52;
OY 42 TDPTFLAHNACTDPPVAVIIGQENPNNSASSEPPREGNNPNQVMDTEHSNPDG 101
DB 118 SDPRDI--DQDNSTSPSI-YSPGSVENDSDSSSLSGCPARPYHPPPLFPSPPPPD-- 172
OY 102 SSVPDPTWGPERRGEESGHLVATGTAA-----GGG 135
DB 173 -SIPROPESGFE-----PHRSVPPTGYHAPMEPTSRLLFGPPPGAPPHPOLYPGSA 224
OY 136 GGLIASPKRG-----ATPLP-----PESTP----- 156
DB 225 GGGVLSGPPMGPRGGAASVGPSPGCKQHPPPTPIPISSSGASGAPPAKPNTPTVGAG 284
OY 157 ----APP-----PPP-----GHUNITLIELRLVLOQ 189
DB 285 NLPASPPAPFPHTVTPMLPPPALPLNNASASPPGMGAQPIFGHLPSP----- 333
OY 190 ROIHOMOMTEOICHOVLLLSLGLQTVG-----APASPELBTGTGASTKPLPLFSP 242

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:17:45 : Search time 43.0643 Seconds
(without alignments)
4808.565 Million cell updates/sec

Title: US-09-988-117-1

Perfect score: 5277

Sequence: 1 MAHSESRSLGVAPGEPAAE.....PSITSTGLSPFRKDDPTIP 1005

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

- 1: SPREMBL.21:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_unclebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1291	24.5	1330	13	Q91929 xenopus lae
2	1280.5	24.3	1261	13	Q902W5 xenopus lae
3	1274.5	24.2	1377	13	Q9DDN5 xenopus lae
4	1246	23.6	1308	13	Q9DF77 gallus gall
5	1176.5	22.3	1272	11	Q9EPW7 mus musculu
6	1145	21.7	1061	13	Q9PVN7 xenopus lae
7	1070	20.3	846	13	Q9PVN6 xenopus lae
8	1038	19.7	1053	4	Q9UJQ4 homo sapien
9	1005	19.0	898	13	O12958 oryzae lat
10	910	17.2	382	5	Q9W7G2 gallus gall
11	784	14.9	1373	13	Q9VKH2 drosophila
12	727	13.8	1263	5	P91639 drosophila
13	726	13.8	1267	5	Q9VKH3 drosophila
14	663.5	12.6	299	13	Q9DGH3 brachydanio
15	648.5	12.3	327	13	Q9DGH1 brachydanio
16	631	12.0	288	13	Q9DGH2 brachydanio

17	513	9.7	549	13	Q9PVN5 xenopus lae
18	391.5	7.4	744	5	Q17396 caenorhabdi
19	363.5	6.9	2406	4	Q9BXS0 homo sapien
20	360.5	6.8	1258	4	Q9P2A7 homo sapien
21	346.5	6.6	2282	11	Q61479 mus musculu
22	342	6.5	751	4	Q9NS43 homo sapien
23	334	6.3	927	4	Q9UEG4 homo sapien
24	333.5	6.3	619	4	Q96K58 homo sapien
25	333.5	6.3	736	5	Q96016 drosophila
26	330.5	6.3	1300	4	O15090 homo sapien
27	329.5	6.2	498	4	Q9H8L4 homo sapien
28	324.5	6.1	725	4	Q96S24 mus musculu
29	324.5	6.1	1237	11	Q9D2D7 mus musculu
30	322	6.1	1615	13	O57415 gallus gall
31	321	6.1	734	4	Q96171 homo sapien
32	316.5	6.0	726	11	Q92A56 mus musculu
33	316.5	6.0	2232	5	P91365 caenorhabdi
34	314.5	6.0	962	5	O61360 drosophila
35	314	6.0	814	11	O70162 mus musculu
36	313	5.9	1920	5	O46205 drosophila
37	311.5	5.9	934	5	Q9VKR1 drosophila
38	308.5	5.8	1893	5	O9W4J1 drosophila
39	308	5.8	1891	5	O77275 drosophila
40	307	5.8	1173	13	Q90783 gallus gall
41	306	5.8	561	4	Q96BV0 homo sapien
42	305.5	5.8	1104	4	Q9P243 mus musculu
43	299.5	5.7	701	11	O99K53 mus musculu
44	299.5	5.7	743	11	O8R5D1 mus musculu
45	299.5	5.7	744	5	Q9VQ09 drosophila
46	298.5	5.6	563	11	Q921H7 mus musculu
47	298	5.6	591	11	O8R0V0 mus musculu
48	298	5.6	841	11	O9WYL8 mus musculu
49	298	5.6	644	4	Q9NT61 mus musculu
50	296	5.6	496	4	Q96SL8 mus musculu
51	295	5.6	744	5	O24219 drosophila
52	293.5	5.6	820	5	Q9Y094 drosophila
53	293	5.6	884	11	Q99PV8 mus musculu
54	293	5.5	291	4	Q96CP9 mus musculu
55	291	5.5	691	4	Q9H937 mus musculu
56	290.5	5.5	1636	4	O9Y474 homo sapien
57	290.5	5.5	596	11	Q9CVF3 mus musculu
58	290	5.5	783	11	Q9QY56 mus musculu
59	288.5	5.5	995	11	O35615 mus musculu
60	288	5.5	650	11	O00146 mus musculu
61	287.5	5.4	784	4	O00146 mus musculu
62	287.5	5.4	662	5	Q9N360 mus musculu
63	287	5.4	891	5	O61361 drosophila
64	287	5.4	835	4	Q9H165 homo sapien
65	285.5	5.4	1190	4	O99621 homo sapien
66	285	5.4	599	4	Q9P2F9 homo sapien
67	284	5.4	613	4	Q9H785 homo sapien
68	282	5.3	754	11	Q99J65 mus musculu
69	282	5.3	624	11	O62065 mus musculu
70	281.5	5.3	786	4	O9COK0 homo sapien
71	281.5	5.3	894	4	O8TDG8 homo sapien
72	280.5	5.3	485	4	O9NRY0 homo sapien
73	279.5	5.3	1173	11	O63624 ratu
74	279.5	5.3	744	11	O62788 ratu
75	279	5.3	618	4	Q96MX3 homo sapien
76	278.5	5.3	703	11	Q9R161 mus musculu
77	278	5.3	568	11	O8R0T2 mus musculu
78	277.5	5.3	594	5	P91805 sarcophaga
79	277.5	5.2	1054	5	O9VA40 drosophila
80	277	5.2	742	4	O9H0M5 homo sapien
81	276.5	5.2	679	4	O8TF50 homo sapien
82	276	5.2	708	11	O91XV1 ratu
83	276	5.2	727	11	O8VC29 mus musculu
84	276	5.2	736	11	O99PJ6 ratu
85	275	5.2	943	5	O8WRV3 drosophila
86	275.5	5.2	556	4	O9UMP5 homo sapien
87	275	5.2	1660	5	O9NKN0 leishmania
88	275	5.2	2703	5	O9VEG7 drosophila
89	275	5.2			


```
OC Archaeosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae:
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=21184122; PubMed=11287198;
RA Farrell E.R., Toth G., Church E., Munsterberg A.E.;
RT "Cloning and expression of CSALZ, a new member of the spalt gene
  family in chick.";
RL Mech. Dev. 102:227-230(2001).
DR EMBL; AF304358; AAK38370.1; -.
DR InterPro; IPR000345; Cytc_heme_bind.
DR InterPro; IPR000822; znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 7.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_7.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
DR DNA-binding; Zinc-finger.
SQ SEQUENCE 1261 AA; 134317 MW; BB22E257BCAE8565 CRC64;

Query Match 24.3%; Score 1280.5; DB 13; Length 1261;
Best Local Similarity 31.2%; Pred. No. 8.1e-67;
Matches 366; Conservative 156; Mismatches 362; Indels 289; Gaps 41;

OY 4 ESESSRLGVPAGEPAELGGDASE-----EDHPVCAKCAOFTDPTREFLAHONACSTDPP 59
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
19 QAEVYSEHAVP-GEGADDGSGNESRSGSEETNVCCEKCAEFKTKDTFLHKKNCTKNPL 77
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
60 VVVIIGQENPNNSASSEPRPEGHNNPQVND--TEHSNPDSSGSVPDPTDPTGPERRGE 117
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
78 VLIIVNDEADAPPAEEPEPSPASSPDQAESEAEQVOPENSESESEVKEEPEPME 137
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
118 ESSGHTLVATGTGAAGGGGGLILASPKIGATPLPESTPAPRPPPPPP----- 165
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
138 VETAEKSFQNOGTSN-----TATPPLPOLPEPSPTMTSYTPMTNTVLTLLST 184
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
166 -----PRPGVSG--HLNIPLIIEELRVLOQROIHQOMTEQICROVLLG 209
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
185 KVAVQFQSARSAASASISGCVAVAIPIMLIEQLMALQOOQIHOQLIQIINSQVAMMN 244
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
210 -----SLGOTVGAPA---SPSELPG-----TGTASSTKLLPLPESP 242
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
245 RQRLRALNLYVVAAPGTGQASNLQGFATSAIQLTAVLPAIMQAAQAQ--PAFDG 302
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
243 IKPVQSKTIASSSSSSSSSGAETPKQA-----PFHIXHPRGSQ 282
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
303 SOHI--SRPTSGASTPNISSSGSSAPRESSAPCSSNATISVTPVSVSNTTISASQPNAS 360
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
283 HPFSAGGVGRSHKPTPAPSPALPGSTDQOLIASPHLAFPTTGLLAOCIGAARGLEMTAS 342
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
361 TPSPSIGH-GLTSTISNLPRLPQT-----SSNSVIFPNPLVSIAM---TANALDPLSA 410
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
343 PGLLKPKNGS-GELSYGEVWMPLEKPGGRHKRCFCAKVFGSDSALQIHLRSHTGERPYKC 401
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
411 --LMKRRKKGKPPNVSVSEPKSSSEDPFFKHKRCFCAKVFGSDSALQIHLRSHTGERPKC 468
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
402 NVGNGNRTTGNLKVHFRHREKYPHYOMNPHRVPHLVLVYTSSGIPGMSVAPREK--- 458
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
469 NIGGNRSTIGNLKVHFRHREKYPHYOMNPHRVPHLVLVYTSSGIPGMSVAPREK--- 528
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
459 --AEEEAATP---GGGVERKPLVASTALSTESTLTLSS-----TSAGATAPGL 503
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
529 TWLDSKRVLLPTIPISIGQLPPTIPGVNSYGDSPSTITPMSRSPQRPASSECTSLSPSL 588
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
504 PAFNKEVLMKAVERK-----NKADE-NTPPGSE-----GSAISGVAESSTATLMQ 547
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
589 NTSSESGVPVSAESQPOVQSSSVTKAEPISTLPASTRLGDHSLSGVSTASTSSIPITVT 648
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
548 LSLMTSLPSAL--LTNHFKSTSGSFPLLCARALGASPESTKLOOLVKEIDGOGAAV 605
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
649 DSSVSTSLPNVLAIVASQCFKA--KFPFGGLDSS--QTSETSKLOOLVENVIDK----- 699
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
606 TSAASGAPTTAPAPSSASSAGSPNOVCICLRVLSCPRALRLHYGHGGERPFCCKVCGR 665
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
700 -----:||||| ||||| ||:| | |||||:|||||
OY 666 ESTRGNLRAHFVGHKASPARAQNSCPICOKKFTNAVTLQOVNRHMLGQIIPNGTALPE 725
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
741 FTTGKNLKHFGVHRAKAPPLVHOSCHPICOKKFTNAVYLOQHIRMHGOIPN--TPLPE 798
OY 726 GGGAA-----QENGSEOS-----TVSAGSFPQOQ 750
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
799 GFODAMDSELSYDEKNTDITLSNFDDDDIDENSMEEPELKDMAKPLISTSGSCF--- 855
OY 751 SQPSPEELSEEE-----EDEEEEDVTDSESLAGC---SES 788
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
856 SSPSPVSISSIALENQMKMIDSVNQCQLTSLKSTIENGSGEHDLSNDSSAVGDLESQS 915
OY 789 GGEKAISVRDSEFASGAEEVGYVAAATACKEND---SNEKTTQOSSL-----PPPPP 841
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
916 AGSPAMS-----ESSSMQALSPVNSNSESFRKSPGLSNOEPEIOLKTEKPDSPPP 969
OY 842 -----DSLID-----QPPMEQSSGVLGKKEGKPRSSSPASALTP-----EGE 882
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
970 ATNGGALDLTSTNPGRPVKEEAPFSLFLNREKPSOSTSPLVTSTAPTMKMEVNGH 1029
OY 883 ATSVTVEELSLQEPAMR-----KEPESSSRKACVCGAAPPQOAL 924
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1030 SKPISLGEVPSLPAGIQVPAAPQTVMSPGITPMLAPPPRPTPROHNCQSCGKTFSSASAL 1089
OY 925 EEHOKTHPKEGPLFTCVFCRQGFLEATLKKHM 957
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1090 QIHERTHTGKRP-FCGCTIGRAFTTKGNLKVHM 1121
```

RESULT 3

```
OY Q9DDN5 PRELIMINARY; PRT; 1377 AA.
ID Q9DDN5;
AC 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DR 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Spalt transcription factor Sal11.
GN SAL11.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Klingbeil P., Frazzetto G., Boumeester T.;
RT "Xsal11, a Xenopus homolog of the human Townes-Brooks syndrome gene
  SAL11.";
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF310007; AAG45108.1; -.
DR HSSP; P15822; IBBQ.
DR InterPro; IPR000345; Cytc_heme_bind.
DR InterPro; IPR000822; znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 10.
DR ProDom; PD000003; znf_C2H2; 1.
DR SMART; SM00355; znf_C2H2; 10.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 9.
DR DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 1377 AA; 148861 MW; 68A9FD22B35A16DB CRC64;
```

```
Query Match 24.2%; Score 1274.5; DB 13; Length 1377;
Best Local Similarity 31.7%; Pred. No. 2e-66;
Matches 359; Conservative 161; Mismatches 336; Indels 277; Gaps 47;
```

```
OY 26 SEEDHPVCAKCAOFTDPTREFLAHONACSTDPPVVIIGQENPNNSASSEPRPEGHN 85
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
114 NKEAH--VCSRCAEFFELSLDLQHKKNCTKNQVLIV---NENPPPAEASFSPPSDN 168
```

Y 86 NQVW-DT-----EH-----SNPDSGSSVPT 106
Y 169 PDEHMDITINNSDQADCSLDSKHGEHGKPDNEEVEVETSTNNSSGSGNKVNDGSSSN 228
Y 107 DPTWG-----PERGESSSGHEFLVAATGTAGCGGGLILASPRLGATPLPPESTPAP 158
Y 229 NPTLGSALTTSPLQVADLTALGNFMSINSVIEN-----QSTVVAVAQFSQEA--- 279
Y 159 PPPPPPPPPGVSGHILNPLLELRYLQORLIHQOMTEQICRQVLLLSIGQTVGAP 218
Y 280 -----KTGGAANNKLAFAFMQGLALQOOQIHLQHLIEQIRHQILLAS--QSAEMP 330
Y 219 ASPSELPGCTASTPRLPLFSPI-KPVQTSKTLASSSSSSSSSG----- 264
Y 331 TSSSDPQAGLRVSAT-PLTTLSSHLSQOLAAAGIQAOSLASQASASIGKMKOLPTIOLPQS 389
Y 265 -----AETPK-QAFHLYHPLGSOH-PEFSAVGGRSHKP-----T 297
Y 330 NSVNTVPSTICSSSPNNMALATYVNPSSERVPSSTGSSQANNQAPAVITSSPPAISLLS 449
Y 298 PAPSPALPGSTDLIASPHLAFSPSTGILAAQCLGAA-----RGLENTASPGI--LKPNG 351
Y 450 PASNPPLLPQSPFN-ATP-----ALPSIGTFAEDLNSLRKSKPVSYAFETKST 499
Y 352 SEGLSYGEVWGLEKQGGKRCRCACAVFGSDALQIHLRSHTGERPRYCNVCGNRPTR 411
Y 500 SDEAFF-----KHKCRCAKAVFGSDALQIHLRSHTGERPRFCNICGNRFSTK 547
Y 412 GNLKHFHRRREKYPHYQMPHPVEHLDVITSSGLPYGMVPPREKABEE----- 462
Y 548 GNLKHFQHKKEYPHIQMPHPVEHLDNISTSGIPTYGMSIPEKPYATNMILDSPILT 607
Y 463 AATPGGVERKPLVASTTALATESITLLSTAGTATAPG--LPAFNKVLMAKAVEPKN 519
Y 608 TLTTSYGMLPRLPTLPRIKTEEPQIPISHTPASPDSVSKSETASLLKKTIDEPD 667
Y 520 KADEMTPPSEGAISGVASSSTRATIMOLSKLMS-----LPSWALL 561
Y 668 EKEAAMP-----LIDKEHQSONSDSLQMLNTPSACSPPTDGSIVMPNILLP--LM 718
Y 562 TNHFKSTGSPPLPLCARALGASP-SETSKLOQLVEKIDROGAVAVTSAASGAPTSAPAP 620
Y 719 SEQFKA--KEPF--GGILDVTPASERTSKLOQVLENDIK----- 753
Y 621 SSSASGPNQCVICLAVLSCPRALRLHYGOHGERPFCVKYCGRAFTSGNLRANFVGHK 660
Y 754 ---SSDPNECVICHVLSQOSALKMHYRTHTGERPFCVKYCGRAFTTGNLKTHSVHR 809
Y 681 ASPARAONSCPTCKKFTNAVTLQOHVBMHLAGOIPNGTA--LPEGGAQOENG-EG 737
Y 810 AMPPLRVQHSCTPCCKKFTNAVVLOQHIRHMGQIPIPTPVAENYPDSSGS--DTGSFDE 867
Y 738 STVSGAGSPPOOQSPSPREEELSEEEDEDEEE-----DVTBDSIAGRSSESG 789
Y 868 KTIDLDLN-SDENMDCPDSSVPDPFKSIDASQDSSSPLPLEVSSITALENOKKLIN 926
Y 790 GEKALSVKRDSEPA-SGAEEVGTAAATAAGKEMDS-----NEKTTQOSSLPP-PP 839
Y 927 AGIALOQLASLSAENGVSVDGMTNDSISLGCDMESQASGPAASESTYSMHALSFPNS 986
Y 840 PPDSDIDOPRMEQSGSYLGCKEKGKPERSSSPASALTP--EGEATSVLVEELSLQE 896
Y 987 TIDYIKSPMTDEKLOAV-----SLDPTNGLSPTPANGALDLT--SSNTDK 1031
Y 897 AMRKEP-----GESSRKACEVCGQAFPSQAALTEHOKTHPKEGFLFTVCFCROGF 947
Y 1032 VIKKEPLVLGPFPRDRCCKYKNTICDICKTFACQSALDIHRSHTKER-FICTVCNRRG 1090
Y 948 LERATLKKHMLLAHOV-----QPPAPHGPNIALSLVPCGSPSITGTLSF 995
Y 1091 STKGMLKOHMLT--HQMRLPLSOLFEP-----SSSWTP--NPTIPSPASNP 1132

RESULT 4
Q9DF77 ID 09DF77 PRELIMINARY; PRT; 1308 AA.
AC 09DF77:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Spalt 1.
GN SALL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20442162; PubMed=10985862;
RA Farrell E.R., Munsterberg A.E.;
RT "Chick spalt 1 gene expression is controlled by a combination of FGF
RT and Wnt signals in developing limb buds."
RL Dev. Biol. 225:447-458(2000).
DR EMBL: AF288697; AAG13011.1; -.
DR HSSP: P15822; 1BBO.
DR InterPro: IPR000345; Cytc_heme_bind.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF00096; zf-C2H2; 10.
DR ProDom: PDO00003; ZnF_C2H2; 1.
DR SMART: SM00355; ZnF_C2H2; 9.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN.1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 9.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 1308 AA; 139846 MW; 81810CC16336298E CRC64;

Query Match 23.6%; Score 1246; DB 13; Length 1308;
Best Local Similarity 30.7%; Pred. No. 8.9e-65;
Matches 352; Conservative 162; Mismatches 360; Indels 272; Gaps 40;

Y 22 GGDASEDHPOVCAKCAQFTDPTFLAHQNAQSTDPVMVTYIGQENPNNSASSEPR- 80
Y 33 GNNTTKNKDAHVGCRCACFEFLSDLLQHKKNCTKNQVLVY--NENASPETTPPRS 89
Y 81 PEGHNPQVWDTEHS-----NPPDSGSSVPTDPTWGPERGESSGHEFLVAAT 128
Y 90 PSDNPDEQMDTYNNMDQVDCSDLSERKNLDEESMDVEASINSSSSSKSVNNSITSS 149
Y 129 GYAAGGGGGLILASPKLG-ATPLPP-----ESTP-APPPPPPPPPVGVSG 173
Y 150 NSTWKTSAVTTSLPHGDLTTLGNFVSYNSVNIENQSTVVAVAQFSQEARCNGASNN 209
Y 174 HUNPLILEELRYLQORLIHQOMTEQICRQVLLLSIGQTVGAPASPSELPGTAST 233
Y 210 KLAVALMEQLALQOOQIHLQHLIEQIRHQILLAS--QNTDMPTSSPSQGLTBASA- 266
Y 234 KPLPLFS-----PIKPYQ-----TKTTLASSSSSSSS 261
Y 267 NPLSTLSSHLSQOLAAAGIQAOSLASQASISGVKQLPPLQIPQSNPGSTLIPSSGSSSP 326
Y 262 S-----SGAETPKQAFHLYHPLGSOHPFSAGVGRSHKP-----TPASP 302
Y 327 NINILAAAYT-----PSSSEKVASISIGSGLSNPVPVSASSSPAPAFISSLSTASNP 377
Y 303 ALPGST--DOLASPHLAFSPSTGILAAQCLGAARGLEATASPGLLKPNGSGEISYGEV 360
Y 378 ILPQPPPSNISFSP-----LSNIGTPAED-----INSLTALAQQRSKPPN---VYAFEA 425
Y 361 MGPLEPGRHKRCRCACAVFGSDALQIHLRSHTGERPRYCNVCGNRPTRRGNLKYNHFR 420
Y 426 KNSDEAFFKHKRCRCACAVFGSDALQIHLRSHTGERPRFCNICGNRSTKGNLKNHFOR 485
Y 421 HREKYPHYQMPHPVEHLDVITSSGLPYGMVPPREKABEEAATPGGVERKPLVASTT 480
Y 486 HKEKYPHYQMPHPVEHLDNIPSTIGIPTYGMSITPPE-----KPV---TS 527

Medline=99458636; PubMed=10527856; Rx Onuma Y., Nishinakamura R., Takahashi S., Yokota T., Asashima M.: "Molecular cloning of a novel xenopus spalt gene (xsal-3)."; Biochem. Biophys. Res. Commun. 264:151-156(1999).
DR EMBL: AB030827; BAA85901.1; -
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf_C2H2; 6.
DR ProDom: PD000003; Znf_C2H2; 1.
DR SMART: SM00355; Znf_C2H2; 5.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 5.
DR DNA-binding: Metal-binding; zinc-finger.
SQ SEQUENCE 826 AA; 89219 MW; B76EB558C9E63D CRC64;
Query Match 20.3%; Score 1070; DB 13; Length 826;
Best Local Similarity 32.4%; Pred. No. 1e-54;
Matches 304; Conservative 131; Mismatches 305; Indels 198; Gaps 35;
OY 18 PAELGGDAS-----EDHPQVAKKCAOFTDPTFLAHQNAQSDPPMVIIGGQENP 71
DB 26 PSSVGVGRVKKRMEETH--ICEKCAEFFELSDLEHKRSCTPTVLIIMDG-EGAM 82
OY 72 NSSASSEPRPEGHNPNQVMDTEHSNPPDSGSSVPTDPTWGERGEESSGHFLVAATGA 131
DB 83 SHDGTTEESPEG-----AAEPSEVPAVDSVQPKRSSVSPKMEKVDNSNKTVPQNSK 136
OY 132 AGGGGGLIASPKLGATPLPEESTPAPPPPPPPPPGVSGHLN---IPLIIEELRVLD 188
DB 137 TGLG--YVPKTMLSMTNVTYLTQINSTKVAANOAHASDVAATNPNAIPILIOGLVCLQ 194
OY 189 QROIHOMQTEQRCROYLLIGSLGQTGAPAS--PSELPGTASTKPLLPFPSPKPV 246
DB 195 QOOLQOITLQEQIRIQIAM-----APNSLAP-----IAATDPLKALGAHLSQ 240
OY 247 QTSKTLASSSSSSSSSGAETPKQAFPHLYPLGSHFSPAGVGRSHKPTPAAPSPALPG 306
DB 241 LSAVAALLIQKAKTQSLSLSKQ-----SKLPHSNVAM-----PTAGTVP-LAL 284
OY 307 STDQLIASPHLAPSTTGLLAOCLGARGLEATASP-GLKPKNSGSLSYEVMGP 365
DB 285 TYSLLKQEBNLGLTNAVGRPPNPALPHSPGTIIIFQNPINALDPSK-KLAKKFTVVTTP 343
OY 366 KRGG-----RHKRCFCVKGVSALQITHLRSHTGERPKYKCNVGNRTGMLKVFHR 420
DB 344 KRGNEQDLRHAKCFCKGKVGNSALQITHLRSHTGERPKYKCNVGNRTGMLKVFHR 403
OY 421 HREKYPVQNPVPRPHLDVITSSGLPYGNSVPERKAEAEATPGGVERKPLVASTT 480
DB 404 HMDKYHIKKNPVPRPHLDVITSSGIPYGMVSPLDESNLADIATKSG-----LT 453
OY 481 ALSATESLTLSTSAGTATAGLPAFNKFLVLMKAVPRKKADENTPTGSGSAISG---V 537
DB 454 GLPSAANLSGLTESV-----LGAFF-----PLN-MOSRPSPGSGESVSSGAVV 495
OY 538 AESSATLMLQSLMTSLPSWALLTNHFKSTGSPFLPLCARALGASP--SETSKLQOLVE 595
DB 496 QESGTD-----QSFNSPPVY-----GSSQSGSETTLQOLVE 527
OY 596 KIDROGAVAVTSAAGATTSAPAPSSASGPNOCVICILRVLSCPRALRLHYOGHGER 655
DB 528 NLDKNG-----SETNECLICHRVLSCPSSLKMHYRHTHGER 563
OY 656 PFCKYVCGRAFTSGNLAHVGHKASPARAONSCPIOCKKTNATLLOOHVMMHGLGO 715
DB 564 PFCKICIGRAFTSGNLTGHVHRANPRLKLOHSCPICOKFTNAVYLOOHIRNMHMGK 623
OY 716 IPIINGTALPEGGAQENGSEOSTVSGA--GSEFPQO-----SQQSP 756
DB 624 IPII-----TPVSEESADLDSMDEKNGELNSTPTDENLDOIIMDEDELAEVAGSKRP 679
OY 757 EEEELSEE-----EEEBDEEBEDVTDDSL-----AGRSGSEGEKALSVAGDSEASGA 806
DB 680 HSETRAESPAWQFSTGTGQDKFVTLPLPALNLOHONSYSKSSSENGSLIEDGLINDSSVMDQ 739

OY 807 EEVGTVAATAAGKEMDSNEKTTQOSSLPPPPPPSDLPQPMQGGSSGYLGGKEGK 866
DB 740 EYPTKSPGOSKRTTSPINSQSDSNAS-KSPSYNGID-----DLGML-----SK 784
OY 867 PERSSPASALTPPEGE-AUSVTLVEELSLQEMARKKEP 903
DB 785 DEHSQN--GSLNDDGQALDLT---NGCFARKIKEEPG 817
RESULT 8
ID 09UJ04 PRELIMINARY; PRT; 1053 AA.
AC 09UJ04;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Djl112P19.1 (novel protein similar to SALL1 (sal (Drosophila))-like 1)
DE (LOC57167)).
GN Djl112P19.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Babbage A.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL034420; CAB61485.1; -
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf_C2H2; 8.
DR ProDom: PD000003; Znf_C2H2; 1.
DR SMART: SM00355; Znf_C2H2; 7.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 7.
DR DNA-binding: Metal-binding; zinc-finger.
KW SEQUENCE 1053 AA; 112231 MW; 61D0D1F21CB2B337 CRC64;
SQ
Query Match 19.7%; Score 1038; DB 4; Length 1053;
Best Local Similarity 29.1%; Pred. No. 1e-52;
Matches 317; Conservative 124; Mismatches 328; Indels 320; Gaps 38;
OY 8 SRLGVAPGEPAELGSD--ASEE-----DHPOVAKKCAOFTDPTFLAHQNAQSD 56
DB 40 AGELGAVVNH-----GNDEVASDEKATVKKLRREETHVCEKCAEFSSIFLEHKKNCTK 96
OY 57 DPPVMVITIGQENP-----NNSASSEPRPEG-----HNNPQVMDTEHSNPPDSGSSV-- 104
DB 97 NPVYL-IMNDEGVPVSEDFSGAVLISHOPTSPGSKCHRENGSSSEDMKEKFDLSEVYL 155
OY 105 -----PTDPTWGERGEESSGHFLVAATGTAAGGGGGLILASPKLGATPLPEESTPA 157
DB 156 KTERALPPTPDISLAKGVANTINTLQA-----LRGTVAVAVNQBSADALPA 203
OY 158 PPPPPPPPPGVSGHLNIPILIEELRVLOQRQHOMQTEQRCROYLLIGSLGQTG 217
DB 204 PVP-----GANSIPVWLEQIILCLOQOLQOITLQEQIRIQVNMNAS----- 244
OY 218 PASSELPGTASTKPLPLPFPSPKIPVQTSKTLASSSSSSSSSGAETPKQAFPHLYH 277
DB 245 ---HALHSAGGA-----DTLKLTHSHMSQOVSAAVA----- 273
OY 278 PLGSHNPSPSAGVGRSHKPTPAAPSPALPGSTDLIASPHLA-----FPSTTGL 326
DB 274 -LLSQKASOGSLDALQAKLPHANIPASATSL--SPGLAPFTLKPDGTVLPRVMSRL 330
OY 327 AAQCLGARGLEATASP-----GLKPKNSGSE---LSYGEVAGLKEKGRHKKFCYK 379
DB 331 PSALLPQAPGSVLFQSPFSTVALDTSKKKGKGRPNISAVVDKPKDEALYVHKCYCSKV 390
OY 380 FGSDALQITHLRSHTGERPKYKCNVGNRTGMLKVFHRHRRKKYPRVQNPVPRPHLD 439
DB 391 FGDSLSLQITHLRSHTGERFPVCSVCGHRTTKGMLKVHFNH-----PQVKANPOLFAEQ 446

```

440 DVTITSSGLPYGMSVPEKAEAEATPGGVERKPLVASTATLSATESLTLSTAGTAT 499
447 DKVAANGIPYALVSVDPIDE-----PSLSIDSKFVLVTTSV----- 483
2Y 500 APGLPAFKFVLMKAVEPKKADENP-----PGSEGS-AISGVASSSTATLMQSK 550
484 --GLQO-----NLSSGTPNPKDLGTGSLPGDLQPGSPSESGPPTLGCYGVNPNP----- 531
2Y 551 LMTSLPSMALTINHFSTGSEFPLICARALGASPSSESKLQOLVETKIDROGAVAVTSAAS 610
532 -----RAGFGSGST-PEP-----GSETLKLQOLVENVNDK----- 560
2Y 611 GAPPTSAAPASSASSGPNOCVITCLRVLSCPRALRLHYGHGGERPFCKYVGRAFSTG 670
561 -----ATDPNECLICHRYLSCOSSLKMHYRTHTGERPFCKICGRAFTSTG 607
2Y 671 NLRAHFVGHKASPARAAQNSCPICQKFTNAVTLQOHVHMLGGQIPNGTALPEGGAA 730
608 NLKTHLGVHRTMTSTIKTQHSPIQCKKFTNAVMLQOHIRMHGGQIPN--PPLPE--NPC 663
2Y 731 QENGSEOSTVSGAGS-----FPQOQSOQSPDEEL----- 760
664 DFTGSEPMTVGNGSTGALCHDVIESIDVEVSSQAPSSSKVPTPLPILHSASPTLG 723
2Y 761 -----SEEEDEDEEBEDVTDEDSLGRGSESGEKALSVRCD 799
724 FAMMASLIDAPGKVPAPFMIQROGSRNGSVESDGLTNDSS-----SLMGD 769
2Y 800 SEASGABEEVGTAAAVATAGKEMDSNEKTQSSLPFRPPDSDLQOPMBQSSGVIG 859
770 QEYOS-----RSPDLETTSFQALSPANSQAESITKSKSP--DAGSKA 809
2Y 860 KKEEGKPE--RSSSPASALTPPEGATSV---TLVEELSLQEAM--RKEPSESSR 908
810 ESSENSRTMEKERSLSLPTFIAPPTYVKEVPGTFVGPSTLSPGMPLLAQPRQAKQ 869
2Y 909 KACVCGGAFPSQALAEHQKTHPKRGPLFTCVFCRGFLERATLKKHMLAHNOVQPPA 968
870 HGTTRCGNFNSASALQIHERHTGKRP-FVCNICGRAFTTKGNLKVH-----YM 918
2Y 969 PHGPONTAA 977
919 THGANNNNA 927

```

```

KW DNA-binding; Metal-binding; Zinc-finger.
FT NON TER 1 1
SQ SEQUENCE 898 AA; 9668 MW; 7B1CF5D7A104287 CRC64;
Query Match 19.0%; Score 1005; DB 13; Length 898;
Best Local Similarity 33.1%; Pred. No. 7,4e-51;
Matches 268; Conservative 102; Mismatches 195; Indels 244; Gaps 28;
2Y 337 LEATAS-----PGLTKPKNS-GELSTGEVWGPLEKPEGGRHKRCQAKVGGSDALQIHL 390
6 IATATAALDPLSALMKHRRKKKPPNVSVFDTKPSSEDPFFKHCRCAKVFSGSDALQIHL 65
2Y 391 RSHTEGEPYKCNWCGRNFTRGNLKVFHHRREKYPHYQOMPHPENHDVYITSSGLPY 450
66 RSHTEGEPYKCNWCGRNFTRGNLKVFHHRREKYPHYQOMPHPENHDVYITSSGLPY 125
2Y 451 GMSVPEKAEAEATPGGVERKPLVASTATLSATESLTLSTAGTATAPGLPA-FNKR 509
126 GMSLPE-----KPV--TWLDSKPVLPVPTSVALQPLPLPSMIGF 167
2Y 510 VLMKAVEPKKADENP-PTPSEGSATIS-GVAESSTAT-----LMQSKL 551
168 AESPLUTPLSRSPQRHSPPSSSECAISLPNVAATDSTRTTSPSPNMLGSDGPLLKPEGI 227
2Y 552 MTLSPSWA-----LTLNHFSTG-----SEF-L 573
228 ILS-PNYSARPENITTTTIVQVLLSTTMTSTSSGSGVSESSISSASNAVSHPLY 286
2Y 574 PLCAALGA-----SPSETSKLQOLVEKIDROGAVAVTSAAGAPTTAPAPSS 622
287 PMLSDQFKAKFPFGGLDMSQTSSETSKLQOLVENVNDK----- 324
2Y 623 SASGNPCVITCLRVLSCPRALRLHYGHGGERPFCKYVGRAFSTRGNLRAHFVGHKAS 682
325 --MTDPNOCVICHRYLSCOSSLKMHYRHTGERPFCKICGRAFTTKGNLKHFEVHNSK 382
2Y 683 PARAAQNSCPICQKFTNAVTLQOHVHMLGGQIPNGTALPEGGAAQENGSEOSTVSG 742
383 PPLRYOHSPIQCKKFTNAVMLQOHIRMHGGQIPN--PPLPE--SLQEMETDL----- 432
2Y 743 AGSPFOQSOQSPDEELSEEEDEEBEDVTDE----- 778
433 --SPDEKSLDAMSNITDDLLDEMEQAMDESDLKEGELDPKPYSPGSSPPTSMISSIAA 490
2Y 779 -----DSLAG-----RGSSESGE-----KAISVRGDSR-----EA 803
491 MENQMKLIDSTANMTGHSFGQKPAQNGSSFGGEADCFITDLSLNAVGDAGSLGALSSES 550
2Y 804 SGABEEV-----GTVAAATAGKEMDSNEKTQSSSLPFRPPD 842
551 SGMQHLSPAHSHSESORSKSPALNNNNSTMTVEBGQENNTAGLTVVKKSEKSETPSPL 610
2Y 843 S-----LDQOPM-----EGSSGYLGKKEB-GKPERSSPASALTPPE--GEAT 884
611 SAIGTGALDLTATPSPKRYIKEESHFSLFLNKDKLSAPNLAISTASNMKMEKNGGK 670
2Y 885 SVTLVEELSLQEAMR-----KEPGESSSKACEVCCQAPPSQALAEHQ 928
671 SLSDNHHLGIQVPAAPATTTMTSPINPMLAPPRPRTPKQHNQOSCGKNFSSASALQIHE 730
2Y 929 KTHPKRGPLFTCVFCRGFLERATLKKH 957
731 RTHGERK-PFACISICGRAFTTKGNLKVH 758

```

```

RESULT 9
112958 PRELIMINARY; PRT; 898 AA.
D 012958;
AC 012958;
JT 01-JUL-1997 (TREMBLrel. 04, Created)
JT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
JT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Zinc finger protein sal (Fragment).
IE Oryzias latipes (Medaka fish).
NC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
NC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
NC Belontiiformes; Atherinichthyidae; Oryziinae; Oryzias.
NC NCBI_TaxId=8090;
N SEQUENCE FROM N.A.
P Stick R., Roester R., Wittbrodt J.;
T "Hedgehog signaling activates spalt at the midbrain-hindbrain boundary
of fish.";
L Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
R EMBL; 077376; AAB51127.1; -.
R HSSP; P15822; 1BBO.
R InterPro; IPR000822; Znf_C2H2.
R Pfam; PF00096; zf-C2H2; 7.
R SMART; SM00355; Znf_C2H2; 7.
R PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
R PROSITE; PS00157; ZINC_FINGER_C2H2_2; 7.

```

```

RESULT 10
09W7G2 PRELIMINARY; PRT; 382 AA.
ID 09W7G2;
AC 09W7G2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Spalt protein (Fragment).

```

3N SAL. *Gallus gallus* (Chicken).
 3C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 3C Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 3C Gallus.
 3C NCBI_TaxID=9031;
 3C (1)
 3C SEQUENCE FROM N.A.
 3C Reveal J.-P., Thaller C., Eichele G.;
 3C "Evidence for morphogenetic signaling by BMP-2 in the chick limb
 3C bud."
 3C Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 3C EMBL: AF110143; AAD3040.1; -
 3C InterPro: IPR000822; Znf.C2H2.
 3C Pfam: PF00096; Zf.C2H2; 5.
 3C SMART: SM00355; Znf.C2H2; 5.
 3C PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
 3C PROSITE: PS50157; ZINC_FINGER_C2H2_2; 5.
 3C DNA-binding; Metal-binding; Zinc-finger.
 3C NON_TER 1
 3C FT 382
 3C FT 382
 3C SEQUENCE 382 AA; 41778 MW; B652263DEFE3B6 CRC64;
 3C
 3C Query Match 17.2%; Score 910; DB 13; Length 382;
 3C Best Local Similarity 47.7%; Pred. No. 9.9e-46;
 3C Matches 194; Conservative 53; Mismatches 92; Indels 68; Gaps
 3C
 3C QY 347 KPKNSGELSYGEVMGRLEKPGGRHKRCFCFAKVFCSDALQTLHRSHTGERPKYKNCGN 406
 3C Db 8 KRPN----VSVTERPKSSSEDPFRFKKRCFCFAKVFCSDALQTLHRSHTGERPKYKNCGN 63
 3C QY 407 RFTTGNLKVHFRHREKXPRHVOMNPRVREHLDYITSSGLPYGKSVPEK----AEE 461
 3C Db 64 RFTSGNLKVHFRHREKXPRHVOMNPRVREHLDYITSSGLPYGKSVPEK----AEE 123
 3C QY 462 EAAP-----GGGVERKPLVASTALSTESTLSTLS-----AG--TPAPGLPAFNK 508
 3C Db 124 KPLVLTPIPSIGLDLPRTIGVNSYGDSPSTTPMKSRPDRPSPASECTLSPLNSES 183
 3C QY 509 FVLKAVERP-----NKADE-NTPPGE-----GSAISGVASSTATLMQSLKM 552
 3C Db 184 GVPVASPPOPVSSSVTKAEPIINLPASTPLGDHSLGCVASTASTSIPTITDSSVS 243
 3C QY 553 TSLPVSAL--LTNHFKSTGSFRLPCALALGASPSPTSKLOLVEKIDRGGAVAVTSAA 610
 3C Db 244 TSLPVPVLAVSQDFKA--KPFEGGLDSM--QTSETSKLOLVEKIDRK----- 289
 3C QY 611 GAPTSAPAPSSASSASNOVCILRYLSCPRALRYLHONGGERPFKCKVCGRAFTSG 670
 3C Db 290 -----MDPRNCVCIGHRLVSCGALKMKHYHTHTGERPFKCKICGFAFTTKG 335
 3C QY 671 NLRAHFGVGHKASPARAONSCPTCKKTTNAVTLQDQVNRHMLGQIP 717
 3C Db 336 NKTHTGVNRHAKRPLRVQHSCTPCCKKFTNAVVLQHRHMHMGQIP 382
 3C
 3C RESULT 11
 3C Q9VKH2 PRELIMINARY; PRT; 1373 AA.
 3C Q9VKH2;
 3C AC 01-MAY-2000 (TEMBLrel. 13, Created)
 3C DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 3C DF 01-MAY-2002 (TEMBLrel. 20, Last annotation update)
 3C DE SARM protein.
 3C GN SARM OR CG6464.
 3C OS Drosophila melanogaster (Fruit fly).
 3C OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 3C OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 3C OC Ephydroidea; Drosophilidae; Drosophila.
 3C NCBI_TaxID=7227;
 3C RN [1]
 3C RP SEQUENCE FROM N.A.
 3C RC STRAIN=BERKELEY;

[illegible]

```

300 -----PSPALPGSTD--OLIASP--LAPSTTGLLAQCSGAARGLEATASPGILKP 348
371 LAKKEMETPRLPFGSLASSITNNHDLPRPNSLDLQKR-----AOEVLDSAGGII-A 425
349 KNGSGELSTGEVWGP---LEKPGGRHKCRFCAKVGSDALSQILHNSHTGERPKCVNGC 405
426 NSMADDFAFGEKSGEKKGRMEPEFKHRCRCYCGKVGFSQSDALQIHNSHTGERPKCVNGC 485
406 NRETTGNLKVHNRHREKRYPHVOMNHRVREN-----LDYVTTSSGLPRGMSVPE 457
486 SRETTGNLKVHNRHREKRYPHVOMNHRVREN-----LDYVTTSSGLPRGMSVPE 545
458 KAEEDAPRRG--GVER--KPLVASTALSTESTLTLSTASGT--ATAPGRAPAKFV 510
546 LGSAPASFPAPRFGILNLYKRPMEILKSLGAARHQYRQDELPTDLKRPQDLEDERQV 605
511 LMKAVERNKADENTPRGSGSALS-----GVAESSTATLMOLSK-----IMTSL 555
606 KNEPEEEDQREHEQEMECSEPEREPRLPLEVRKKEKVEQEQEKQEDHRIEPRRTPS 665
556 PSMALLT--NHFKSTGSPR--PLICARAL--GASRSETSKLOQ----- 593
666 PSSEHRSFNHHHSHNGYRVVYVPIORALIMHPOSSQSHLDHLPTPGQLPRDEFEA 725
594 -----VEKIDROGAVALVTSASGAPTTSAAPR----- 620
726 ERFPLNFTAKMLSEPHNSFVRSRAGALPRGVPRPHHNMARSPEFNKIKHEAAL 785
621 -----SSSASGPNOCYICLRVLSCPRALRYLHYQ 650
786 LPRHNSNDSMNFLEVSNTCEMFKLKLKELMNKKTISDPOCVCORVLSCKSKALOMIYRT 845
651 HGEREPKCKVCYSGRAFSTRGNLKAHNVGNKASPARAKONSQCIQKFFNAVTLQOHVMA 710
846 HTGEREPKCKVCYSGRAFSTRGNLKAHNVGNKASPARAKONSQCIQKFFNAVTLQOHVMA 905
711 HLG-----GQIPNGGTALPRGGGAQDE 732
906 HTGERPTDLPDIOAETRDPRPSMGRNEMNPRMAAFHNGALP--GGPGPRPMNGAH 964
733 NGSEOSTVSGAGSFPOQSOQSPREELISE-----EEREDEEEDD----- 774
965 NGALGSE--SSGGMDMDNDGCGEDYDDVSEHLSNSMLEQJGDSRSRSGDSEKSLFEOKL 1023
775 -----VYDEDSLARGSESGEKA---ISVRGDE---EASGAEVYG--- 811
1024 RIDATGVVNTNPRVPRSSASSHGHVSGTSAPTSIVHASQVILKRSSPARSEASQAL 1083
812 --TVAATAATAGKEMDS--NEKTTOOSSLP-----PPRPDLSL----- 845
1084 DLPTRAPARTSSSSSRSPRLKPKRVPSPSLRSPSSGSSHASANLITSLRPLRVGIDCLPRG 1143
846 -----QOPRMEGSSGVLG----- 860
1144 LQNHLLQOONHLMQOQAAVAAAHAAQHNNHQAALQHOEQLREAAEOKAAAAA 1203
861 -----KEBGG-----KPRSSSPASALAPR-----GEATSVTLVEE 891
1204 AAAAAAAROTPRROARDOROBEGRGAGRPNNPLMGARPRGFMFNLPPLFPRATTONMCA 1263
892 LS--LQEAMKRP-----GESSSRKACEVCGQARPSOALEHOKTHNREGLPTCYF 942
1264 MNOIASVMAAPRNPLALSGVRGS--TTCGICVKTFFCSHLETHNHSHTERP--FKCSI 1321
943 CROGFLERATLKHKMLLANHVOVQPARPHQROPIALSLVDCSPSITSTGSPFR 997
1322 CDRGFTTKGMLKQHMILT--HKIRDMEOETFRNRAY-----KYVSLPISPIR 1365

```

SUPT 12
1639
P91639 PRELIMINARY; PRT; 1263 AA.

```

AC p91639;
DT 01-MAY-1997 (TREMBLrel, 03, Created)
DT 01-MAY-1997 (TREMBLrel, 03, last sequence update)
DT 01-MAR-2002 (TREMBLrel, 20, last annotation update)
DE Zinc-finger protein SALR (SPALT-related protein).
GN SALR OR CG4881.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RA Barrio R., Shea M.J., Garulli J., Lipkow K., Gaul U., Frommer G.,
RA Schuh R., Jackle H., Kafatos F.C.;
RT "The spalt-related gene of Drosophila melanogaster is a member of an
RT ancient gene family, defined by the adjacent, region-specific homeotic
RT gene spalt."
RL Dev. Genes Evol. 206:315-325(1996).
CC 1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
EMBL: Y07653; CA68937.1; -.
DR HSSP; P15822; IBO.
DR FlyBase; FBgn0000287; salr.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00355; ZnF_C2H2; 8.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ
SEQUENCE 1263 AA; 138991 MW; 2025B3BE7417622 CRC64;

Query Match 13.8%; Score 727; DB 5; Length 1263;
Best local similarity 23.5%; Pred. No. 2,1e-34;
Matches 275; Conservative 144; Mismatches 397; Indels 356; Gaps 39;

QY 67 QENPNNSASSSEPRREGNNPQVMDTEHNSNPRDSSGVPPDPTGPERKGESSGHF--- 123
DB 38 REHDRTLLAKELDADSDNSNGTEPQMEAAVAPESDTEREAEREGERDEQDEPENSNALDLS 97
QY 124 LVATGTAAGGGGLI-----LASPKLGATPRPESTPAPRPPPPPPPPPPVSGSH 174
DB 98 LISSGREGSLPGSHVSLDALQHTKVAQAQAFATMAGNQSA----- 140
QY 175 LNIPLILEELRVLQROTIHQMTQICROY-----LLLG--SLGQTVGAPASSELPGT 227
DB 141 -DIAMVOSTIFNVORHIMQLOLIQHSQILKRAEAAALGRHSHSDEEERPERPERPK 199
QY 228 GTASSTKPLLPFLSPKRVQTSKTLASSSSSSSSSSSGAETPKQAFPHLHNLGSHPS- 286
DB 200 QPTNGLKEELELEGGPSESDHES--RRENSKTDKRGTEBRKA-----EEGYQSMCD 252
QY 287 -----AGVGRSHKTRAPS-----PALPGSTDOLIASPHLAPSTTGLLAQCSGAARG 336
DB 253 ISSSLASSITNNHDPAPRNPNCLMQLRREEVLD-----ASQSLHAAQ- 299
QY 337 LEATAPRGLKPKNGSELSTGEVWGPRLKPGGRHKCRFCAKVGSDALSQILHNSHTGE 396
DB 300 MOEEYSEYASKEAOSREIF-----KHCKKVCGLFGYSALQIHLSHTGE 346
QY 397 RPYKCVNCGNRETTGNLKVHNRHREKRYPHVOMNHRVRENLDYVTTSSG----- 447
DB 347 RFYVCVNCGSKFTTKGMLKHNQKHTQIFRPMILPRGVARN-----VGHSGOGVOGBOY 401
QY 448 ---LRYGMSVPRPEKA-----EEAATPGGVGRKPLVASTALSATSESLTLSTSA 495
DB 402 PIRLPFAPVAVPVGQEOHONQVEERPEIRVQAEDLSKRMVKEKE--KSHSPVCAVTPK 459
QY 496 GTATAAPGLPAFNKFLVLMKANVERPKKADENTPRGS-----EGSAISGVASSTATL- 545

```


347 RPFVNCVCGSKFTTKGNLKVHYORHTOIFPMLLPBGVAPN-----VGHSGQGVQVEQY 401
 448 ---LPYGSV-----PPEKAEEBAATPGGVERKPLVASTALSAATESLTL 491
 402 PIRLPFAFPVAVGVGROHONQVEERPEIRQELIPVQADELSKPMVEKE--KSHSPERV 459
 492 STSAGTAAPAGLPAPFK--FVLMKAVER-----KNKADENTPPGSE-----530
 460 KTPKEVKDAAIPSSSKPEKEIKPVYSSRRNGSVRRKQTSVSPQEDREDRLVEHL 519
 531 -----GSAISGVAESSTALTMOLSKLMTSLPSMALLTNHFRSTGSPFLPLCARALGASP 584
 520 IAKLVRRSSASSESPAEVSLAOKERIIDK--SWEDLIEIDKT-----560
 585 SETSKLOOLVERKIDRQGAVAVTSASGAPTTASAPSSASSGPNQCVICLRVLSCPRAL 644
 561 SETSKLOOLVDNIENK-----LTDPNQICICQKMSCHSSL 596
 645 RLHVGQHGGERFKKVCGRASTRGNLRAHFVGHKASPARAONSCPTCKKFTNAVTL 704
 597 QMHRTRTHGERPFCKICGRAPATKGNLKAHMSIHKIKRPMRSQFCPCVCHOKFNGIIL 656
 705 QOHVYRML-----GGO-IPNGCTALPGGGAOENG-----EOS 738
 657 QOHIRIHTMDOSGGOGAPAAANGEAERLGIEDONSNSLGTSTLIDBSTTTSDHSGQRS 716
 739 TVSAGSPPOOQSOQ-----PSPEELSEEEDEDEEDVDYDEDS---780
 717 ESSGCGPDEFMTDSTDSDRNSAATATPHPLERDRERERERRIPNDSDESHSN 776
 781 --LAGRSESGEKALISVGDSEASGAEEVGVAAATAGKENDSNEKTTQOSSLPP 838
 777 PDLTGSESGEMPRAMDLSPPSSN-SGRIFATGLANGAAGG---SGNGGLPMLCMPMP 831
 839 P-----PPDSL-----844
 832 PMLLMAAREEMHNLGNHAKFPLLPFGPLGFMGLRPPVNCNLCFKMLPSLALFSL 891
 845 -----DOPRMEGSSGVLGK-----EEGK-PRSSS 872
 892 QSEHAKBRATGHAORPQSDAGSP-YGAKLTLPNLFAKKPPSSSSSSGKLPRESSNP 949
 873 PASALTP-----EGEATSVTL-----888
 950 PPAENPAPATPKEDPDQOLMVEGASAGESSGTGATSNYPQAGDAEQSLMKQMLNAN 1009
 889 ---VEELSLQEA-MKREGEES-----SRKACEVCGAPFSSQALFEHQKTHPKGPPLF 938
 1010 RFPASPLDFQOALMSAGPRTSLDPVNNKHFCHVCRNFTSSSSALQIHMRTHGDKP-F 1068
 939 TCVCRCRGFLERATLKMLMLNHOVQFAPRHPONTAIALSLVPG 983
 1069 QCNVCAKFTTKGNLKVH-GTHMTNPTSRGRMSLLEPMRPG 1112
 RESULT 14
 DG33
 O9DGH3 PRELIMINARY: PRT: 299 AA.
 O9DGH3: 01-MAR-2001 (TREMBLrel. 16, Created)
 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 Putative spalt protein (Fragment).
 SPAL.
 Brachydanio rerio (Zebrafish) (Zebra danio).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 Cyprinidae; Danio.
 NCBI_TaxID=7955;
 [1]
 SEQUENCE FROM N.A.
 Camp E.M., Lardelli M.T.;
 "PCR amplification of spalt exon sequences from various

RT vertebrates.*;
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ293862; CAC05355.1; -.
 DR InterPro: IPR000345; Cytc_heme_bind.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; Zf_C2H2; 3.
 DR SMART: SM00355; Znf_C2H2; 4.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_2; 4.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_1; 1.
 DR DNA-binding; Metal-binding; Zinc-finger.
 FT NON_TER 1
 FT TER 299
 SO SEQUENCE 299 AA; 33064 MW; A9D03F130DA5FAB5 CRC64;
 Query Match 12.6%; Score 663.5; DB 13; Length 299;
 Best Local Similarity 42.4%; Pred. No. 2e-31;
 Matches 150; Conservative 38; Mismatches 63; Indels 103; Gaps 12;
 QY 404 CGNFTTNGNLKVFHRRHEKYPHVQMNPPVPEHLDYVITSSGLPYGMSVPEKAEEBA 463
 DB 1 CGNFTSTGNLKVHFORIKEKTPHIQMNPPVPEHLDNIPSTGLPYMSMP-----54
 QY 464 ATPGGVERKPLVASTALSAATESLTLSTAG--TATAPGLPAF-----NKPYLMKAV 515
 DB 55 -----KPY--TSWLDKRPVLSTLSSVGMPLPPTIPSLPFIKEENNLSAISP 102
 QY 516 EPKKADE--NTP-----PGSEG-----SAISGVAS 540
 DB 103 SHSAKSDSGPADTPPKNDVLEEGESTLPTNSGAENNSNLSMMSAVGTTIEY 162
 QY 541 STATIMOLSKIMTSLPSWALLTNHFKSTGSFPL-----PICARALGASPSFETSKLOOLVE 595
 DB 163 TTSN-----SPMATNPMLPIMSEQFKA--KFPFGGLDPL-----QGETSKLOOLVE 209
 QY 596 KIDRQGAVAVTSASGAPTTASAPSSASSGPNQCVICLRVLSCPRALRHYGQGER 655
 DB 210 NIDRKVA-----DENECVICHRIISQSLKKNHYRHTGR 245
 QY 656 PFKCKVGRAPSTRGNLRAHFVGHKASPARAONSCPTCKKFTNAVTLQOHV 709
 DB 246 PFKCKVGRAPSTTKGNLKVH-SVHRAMPRLRVQHSCTPCKKFTNAVTLQOHV 299
 RESULT 15
 Q9DGH1
 ID O9DGH1 PRELIMINARY: PRT: 327 AA.
 AC 09DGH1: 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative spalt protein (Fragment).
 GN SPAL.
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OC NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Camp E.M., Lardelli M.T.;
 RT "PCR amplification of spalt exon sequences from various
 RT vertebrates.*";
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ293864; CAC05357.1; -.
 DR InterPro: IPR000345; Cytc_heme_bind.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; Zf_C2H2; 3.
 DR SMART: SM00355; Znf_C2H2; 4.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 4.
 DR DNA-binding; Metal-binding; Zinc-finger.

```
FT  NON_TER      1      1
SQ  SEQUENCE      327 AA: 35908 MW: 1623D9E63F7E9689 CRC64:
Query Match      12.3%; Score 648.5; DB 13; Length 327;
Best Local Similarity 40.2%; Pred. No. 1,7e-30;
Matches 145; Conservative 46; Mismatches 81; Indels 89; Gaps 10;

OY  404 CGNRFTTRGNLKVHFRHRRKRYTHVQNNPHRVPPEHLVDYITSSGLPYGMSVPEKAE64
DB  1 CGNRFTTRGNLKVHFRHRRKRYTHVQNNPHRVPPEHLVDYITSSGLPYGMSVPEKAE64
OY  464 ATPGGVERKPL---VASTALSTESLTLLSTAGTATA-----PGLPAF---- 506
DB  60 ----WLDKRPILPTITSAVLQLPPTIPSTIGSYGDSSTSPRLKRSPPRPSPSECTS 114
OY  507 ---NKFVLKAV-----EPKNAKADENTPPGSESAI-----SGVAESSTATLMOL- 548
DB  115 LSPNHLITETSAIQSISSPPQNLASNTPPVYLKPEALHLPTNSTRGETSISTASISQVI 174
OY  549 -----SKLMTSLPSMALLTNHFKSTGSPFLPLCARALGASSETS 588
DB  175 STTIVTTCSTRTQLTDPVNSSAVSHPSLSQISSNQFNP--KPEGGLDSM--QTSETS 230
OY  589 KLOQLVERIDROGAVAVTSAASGAPTTSAAPASSASGPNOCVICLRVLSCPRALRLHY 648
DB  231 KLOQLVENIDRK-----MTDPNOCVICLRVLSCOSALMKMT 266
OY  649 GONGERPEPKCKVCGRAFTSGNLRRAHFVGHKASPAARAQNSCPICQKKFTNAVTLQOHV 708
DB  267 RHNGERPEPKCKICGRAFTTKGNLTKHFVGHRSKRPFLVQHSHCPICQKKFTNAVTLQOHV 326
OY  709 R 709
DB  327 R 327

RESULT 16
OY  09DGH2 PRELIMINARY; PRT; 288 AA.
AC  09DGH2:
DT  01-MAR-2001 (TREMBLrel. 16, Created)
DT  01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT  01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE  Putative spalt protein (Fragment).
GN  SPALB.
OS  Brechyanio rerio (Zebrafish) (Zebra danio).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC  Cyprinidae; Danio.
OX  NCBI_TaxID=7955;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Camp E.M., Lardelli M.T.;
RT  "PCR amplification of spalt exonlc sequences from various
RT  vertebrates."
RL  Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AJ293863; CAC05356.1; -
DR  InterPro; IPR000345; CytC_heme_bind.
DR  InterPro; IPR000822; Znf_C2H2.
DR  Pfam; PF00096; zf-C2H2; 2.
DR  SMART; SM00355; Znf_C2H2; 4.
DR  PROSITE; PS00190; CITOCHROME_C; UNKNOWN_1.
DR  PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR  PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW  DNA-binding; Metal-binding; Zinc-finger.
FT  NON_TER      1
FT  SEQUENCE     288 AA: 32261 MW: A07A1F2BBD6A6F42 CRC64:

Query Match      12.0%; Score 631; DB 13; Length 288;
Best Local Similarity 41.8%; Pred. No. 1.5e-29;
Matches 143; Conservative 40; Mismatches 67; Indels 92; Gaps 9;
```

```
OY  404 CGNRFTTRGNLKVHFRHRRKRYTHVQNNPHRVPPEHLVDYITSSGLPYGMSVPEKAE64
DB  1 CGNRFTTRGNLKVHFRHRRKRYTHVQNNPHRVPPEHLVDYITSSGLPYGMSVPEKAE64
OY  464 ATPGGVERKPLVASTALSTESLTLLSTAGTATAPGLPAFNKFLKAVPPKNADE 523
DB  56 -PPLSWLDSKPLVGNISIGFMSSSL-----PGLPV----IIRKEEGVSTIKP 98
OY  524 NPPGSE-GSAIGVADS-----STATLMOLSK----- 550
DB  99 HSPVSEELTRKINGHOGFVCSPLIISNEKFOEVNQRLSVSTLRSRSGREDIAINTS 158
OY  551 ----LMTSLPSMALLTNHFKSTGSPFLPLCARALGASSETS KLOQLVERIDROGAVAVT 606
DB  159 VNTGLTKIKSEGLEAKFL--LGSLPNPLGA-----SETSKLEQVENIDRK----- 203
OY  607 SAASGAPTTSAAPASSASGPNOCVICLRVLSCPRALRLHYGONGERPEPKCKVCGRAF 666
DB  204 -----YDPNKGICLRVLSCOSALRMHRTHTGERPEPKCKVCGRAF 245
OY  667 STFGNLRRAHFVGHKASPAARAQNSCPICQKKFTNAVTLQOHV 708
DB  246 TTKGNLTKHTVSIHRSMPLRIQNSCPICQKKFTNAVTLQOHV 287

RESULT 17
OY  09PVN5 PRELIMINARY; PRT; 549 AA.
AC  09PVN5:
DT  01-MAY-2000 (TREMBLrel. 13, Created)
DT  01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT  01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE  Xsal-3', long form (Fragment).
GN  Xsal-3'.
OS  xenopus laevis (African clawed frog).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC  Xenopodinae; Xenopus.
OX  NCBI_TaxID=8335;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  MEDLINE-99458636; Pubmed-10527856;
RA  Onuma Y., Nishinakamura R., Takahashi S., Yokota T., Asashima M.;
RT  "Molecular cloning of a novel Xenopus spalt gene (Xsal-3).";
RL  Biochem. Biophys. Res. Commun. 264:151-156(1999).
DR  EMBL; AB030826; BAA85902.1; -
DR  HSSP; P15822; 1BBO.
DR  InterPro; IPR000822; Znf_C2H2.
DR  Pfam; PF00096; zf-C2H2; 5.
DR  PRODOM; PD000003; Znf_C2H2; 1.
DR  SMART; SM00355; Znf_C2H2; 1.
DR  PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR  PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
KW  DNA-binding; Metal-binding; Zinc-finger.
FT  NON_TER      1
FT  SEQUENCE     549 AA: 59348 MW: ECBBA478D0F088A46 CRC64:

Query Match      9.7%; Score 513; DB 13; Length 549;
Best Local Similarity 30.5%; Pred. No. 2.7e-22;
Matches 134; Conservative 60; Mismatches 142; Indels 98; Gaps 13;

OY  585 SETSKLOQLVERIDROGAVAVTSAASGAPTTSAAPASSASGPNOCVICLRVLSCPRAL 644
DB  5 SETSKLOQLVENIDRK-----SETNCLDICHRYLSCPSL 40
OY  645 RLHYGONGERPEPKCKVCGRAFTSGNLRRAHFVGHKASPAARAQNSCPICQKKFTNAVTL 704
DB  41 KMHRTHTGERPEPKCKICGRAFTSKNLTKHFVGHRAHPPLKQSHCPICQKKFTNAVTL 100
OY  705 QOHVRHLAGQIPNGGATLPEGGGAQENGSEOSTVSGA--GSFPQOQ----- 750
DB  101 QOHIRHMGCKIRP---TPVSEASDDIDSMMDEKNGELNLSFTDENLDDIMDEDELA 156
```

751 -----SQQPSPEELSEE-----EEDEEEEDVYDEDSL-----AGRGSESGEKAIS 795
157 ENASGSGPPTPHSETRESPAMQSTGTGQDKPVTLPALNTQROMSVSENGSLSDG 216
796 VAGDESEASGAEEVEGVVAAATAGKEMDSNEKTTQOSSLPPPP-----PPDSL 845
217 LINDSSVWDQEXPTGKSPQSEARTSPPTSOSDSMSKSPSYNGLDLGLMSKDEHS 276
846 QOPPMQSGSSGV-----GG-----KEEGKPEERS-----SSPASALPPEGATSV 886
277 QNGSLNPDDGALDLITNGGFAARKIKEEPGLHONGEFGRLPNLYGAPPALI--KMEVSSD 334
887 TLVEELSLOEAMKREP-----ESSSRKACEVCGQAPPSQALAEHQTHPKKGPL 937
335 RMAAGATQYIGRPMLSPGLNPLIVPQRSASAKHICTMCGKNFSASALQHERHTGEBK 393
938 FPCVCRQGFLEATYTKKH 957
394 FACTICGRAFTTKGNLKVHV 413

RESULT 18

PRELIMINARY; PRT; 744 AA.

Q17396 Q17396; Q19483; Q17397; 01, Created)
Q17396; Q19483; Q17397; 01, Created)
01-NOV-1996 (TREMBLrel. 01, last sequence update)
01-NOV-1996 (TREMBLrel. 01, last sequence update)
01-JUN-2002 (TREMBLrel. 21, last annotation update)
SEM-4 protein.
SEM-4.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
Rhabditidae; Peloderinae; Caenorhabditis.
NCBI_TaxID=6239;

SEQUENCE FROM N.A.
STRAIN-BRISTOL N2;
Baason M., Horvitz H.R.;
Genes Dev. 0:0-0(1998).
[1]
SEQUENCE FROM N.A.
Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
- SUBCELLULAR LOCATION: NUCLEAR (By similarity).
EMBL; Z71260; CAA95798.1; -
EMBL; U60112; AAB03333.1; -
EMBL; U60113; AAB03334.1; -
TRANSPAC; T02278; -
InterPro; IPR000822; Znf_C2H2.
Pfam; PF00096; zf-C2H2; 7.
PRINTS; PR00048; ZINC_FINGER.
ProDom; PD000003; Znf_C2H2; 1.
SMART; SM00355; Znf_C2H2; 7.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
Alternative splicing; DNA-binding; Metal-binding; Nuclear protein;
Zinc-finger.

SEQUENCE 744 AA; 81694 MW; BCF468C639A30182 CRC64;
Query Match 7.4%; Score 391.5; DB 5; Length 744;
Best Local Similarity 21.8%; Pred. No. 5e-15;
Matches 194; Conservative 89; Mismatches 278; Indels 327; Gaps 32;

219 ASPSELPGTGTASTKPLPLPLFSPKIPVQTSKTLASSSSSSSSSGAETPKQAFHLHP 278
18 SKPRMSSGGDA-----MMSPI-----DLSTKSDENNCER-----48
279 LGSQHPFSAAGVGRSHKPPAPSPALPGSTDOLIASPHLAPPST-----GLIAOCLGA 333
49 -----GAGG-----ALP-LIEDRSNILPHFSVFANPQQLSLCAQGNSS 87
334 ARGLEATASPLGLKPKNGSGELSYGEVMGPLEKPGGRHKRCARCAYVGSDSALQIH-LRS 392

88 SRVVSSTAATSSCP-----IQOSQSFSSPALTLWHLDA 123
393 HTEGRP-YKCNVNCNRRPTTRGNLKVHFHRRKRYHYVGNPHVRLDVLITSSGLPYG 451
124 HEDEQLFSCVCTTTS-----NGODIRH-----149
452 MSVPEKAEEMAPPGGVERKPLVASTTALSTATSLTLLSTAGATAPGLPAFNKFLV 511
150 -----KQKTLASRMSVSPSTIPSSVCFLS-----TPTRPCL-----QFSI 186
512 MKAVEPKKADENTPPGSEGAISGVAESSTATTMLQSLMTSLPSWALLTN----FKS 567
187 NESIGTSEIEEDEEDMDVDGEHVANQLFGHLLQKSDCKSKMAS--LFNHAFFPFAA 243
568 TGSFPLPLCAR-----ALG-----ASPETSLOOLYEKIDRGAV 603
244 FPMMPPEFLMRQPPDPRAVDFAAGHNDMDWEALMETISTSDAEKTRALV-----GDK 297
604 AVTSAASGAPTTAPAPSSASGPNOCVICLRVLSCPALRLHYGOHGGEPKFCVCG 663
298 AV-----PTTD-----PNCILCRVLSCSKALQMHYTHGERFCKICQ 339
664 RASTRGNLRAHFVGRK-----AS 682
340 RAFTTKGNLKTGMGVHRSKHSFRCGLPLISLPQLAAMHQHQAAPPQRTHINPPTSAM 399
683 PAA-----RAONSCIPKPTNAVTLOQHVBMHLGG-----QIPNGGTA-----LP 724
400 AANAVAOIASOCCPLCOQRLNMGELAVHITEHNSLTOPRAWPTPTTTRVOTPFVP 459
725 EGGGAQENGSEOST-VSAGSPFOQSOQPSPEELSEEDEEEDVYDEDSL 783
460 FETTPSLNATDMSTQNNLNLITSAQLKNDSSPNMTDSVEEKITRD-----DPRKMAS 513
784 RGSSESGEKAISVRGBDEASGAEEVGTVAATAAGKEMDSNEKTTQOSSLPPPPDS 843
514 LSPNSSDSSSVRODILSESSEFEKT-----KLEPPILQGVSTPNP---559
844 LDQPMQEGSSGVLGKEEGKPEERSSPASALPPEGATVTLVEELSLOAMKREP 903
560 -----KKNPPLAMKMAETEPPIPRQMPV-----585
904 ESSSRKACEVCGQAPPSQALAEHQTHPKKGPLETCVFCRGGLERATLKKHMLAHQ 963
586 -LSKHQCGVCFKHFSSSALQIMHRTTGDKP-FKCDMCGRAFTTGNLKVHMGTHSQ 642
964 VQPF-----APHQPNTAALSLVPCSP--STISGLS 994
643 QSPSRGRRIEDVAVSVYTERPMLQSPILPTSGAPGASPLAMLGPNGLS 690

RESULT 19

PRELIMINARY; PRT; 2406 AA.

Q9BZS0 Q9BZS0; 01-JUN-2001 (TREMBLrel. 17, Created)
Q9BZS0; 01-JUN-2001 (TREMBLrel. 17, last sequence update)
01-JUN-2001 (TREMBLrel. 21, last annotation update)
01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE kappa B and V(D)J recombination signal sequences binding protein.
GN KRC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RP TISSUE=BRAIN, AND THYMOCYTES;
RC MEDLINE=21100880; PubMed=11161801;
RX Hicar M.D., Liu Y., Allen C.E., Wu L.C.;
RT "Structure of the Human Zinc Finger Protein HIVEP3: Molecular Cloning,
Expression, Exon-Intron Structure, and Comparison with Paralogous
Genes HIVEP1 and HIVEP2.";


```
250 HHPQVLAQOGS-----GSSPKATDIPASAPPEVAVPFVKOSPGHOSPLASPRVPCQP 304
243 IK-----PVQNSKTLASSSSSSSSGAE-----TPKQAFHLYHPLGQHPR-- 285
305 LKEEDDEGPVDS---SPSSQSPSSGAEADSDNSPASS-----SSRPLKV 351
286 -----SAGVGR-----SHKPTAP-----SPALP----- 305
352 RIKTIKTSQGNIRTYQVSPDPDPAPLACAGFLAASLIKSPATPTSEGPVVSQVL 411
306 -----GSTDOLIA-----SPHLAFTSTGLAACL 331
412 GDGTRIKGTVLPVATTIONASTAMLAASVARKAVVLPGGTATSPKMAKIVLGIY-PCAL 470
332 GAARGLEATASPG-----LLPKNGSGEL----- 355
471 PKADGAGAGTGGOKVNGASVWVQSPKATGPTGGTGVISHTOSSLYEAFNKILNSKN 530
356 ---SYGEVNGP-----LEKPGRHKRCFCAKVFGSDSALQIHLSRHTGERPYKCNVGNR 407
531 LLPATRPNLSPPRAEAGLALPPTGYRCLECGDAFLEKSLARHYDRSMRELVTCNHCAAR 590
408 --FTTRGNLKVFHRRREK-----YPHVQNPHPVPEHLDYVITSSGLPYGMSVPERKAE 460
591 LVFFPKSCILLHAREHKDGLVMQCSHLMRPVALDQWVGQPDITPLP--VAVVP--VS 646
461 EEAATPGGVERKPLVASTALSTALSTESLT--LTSAGTATAPLPAFNKVLMAKAVEKN 519
647 GPLALPALKGEGATISSAITTYVAEAPVLPSTE---PPAAPTASATCFCEKLEKE 701
520 KADENTPPESEGAISGVASTATLMQLSKIMTSLPSWALLTNHKSSTGSPFLPLCARA 579
702 QCRDAGMAAHFQQLGPPRAGATSNVCTCPMM--LPNCSFSAHQRMHKNRPHVCEC 759
580 LGASSETKLOLVKIDQOGAVAVTSAAGAPTTSAAPSSASAGNOCVYICLRYLS 639
760 GG-----NFLQANFQTHLEACLHVSRRVG-----YRCPSCSYVFG 795
640 CPRALRLHNGOGERFERFKVCGRFASTRCGLRAHFVGHKASPAARAONS-----CPIC 694
796 GVNSTKSHQTSICEVFHKPCICPMAFSGPSAHNL--YSHFSTQTOQAKLITKCAWC 853
695 QKFTNAVTLQOQHVHMLGQOIPNGGALPBGGAQAENGSEOSTVSGAGSFPOQSOQP 754
854 DTVFHKRPLLSHFQDL-----LP-----QGVSVFKCPSCLLFAQKR 892
755 SPEELSEEEDEDEEDVTDSDSLAGRSESGEKAISVAGDSEASGAEVEGIVA 814
893 TWLEHLKNTHQSGLEE-----TAGKA--GGALLTPKTEPEELVVSQ-----G 934
815 AAATAGKEMDSNKTTOOSSLP PPPPSLDOPOMEGSSGVLGCKEKGK----- 866
935 GAAPRTSESSSSSEEEVPSPPRRP--AKRPR-RELGSKILKG--GGPGGWTGCLC 989
867 ---DERSSSPASALPPEGATSVTLVEELSLQEAARKPEGSSSSRKACEVCGOAPSOA 922
990 HSMFPERDEYVAH-----MKKEHGKSVKKFPCRLCERSFCSAP 1027
923 ALEENOK-THPKEGPLTYVFCROG---FLERATLKKNHLLAH 961
1028 SLRRHVNRHGGIKRVPKRYCTEGKRTSSRLLEKHQVHR 1070
```

```
ESULT 21
61479
D 061479 PRELIMINARY; PRT: 2282 AA.
C 061479;
T 01-NOV-1996 (Tremblrel. 01, Created)
T 01-NOV-1996 (Tremblrel. 01, Last sequence update)
T 01-MAR-2002 (Tremblrel. 20, Last annotation update)
E DNA binding protein RC.
N KRC OR RC.
```

```
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=THYMOCYTE;
RX MEDLINE=97001141; PubMed=8812474;
RA Wu L.C., Liu Y., Strandmann J., Mak C.H., Lee B., Li Z., Yu C.Y.,
RT "The mouse DNA binding protein Rc for the kappa B motif of
RT transcription and for the V(D)J recombination signal sequences
RT contains composite DNA-protein interaction domains and belongs to a
RT new family of large transcriptional proteins."
RL Genomics 35:415-424(1996).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: L46815; AAA0884.1; -.
DR HSP; F15822; IBB0.
DR MGD; MGI:106589; Krc.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 5.
DR PRINTS: PR00048; ZINC_FINGER.
DR SMART: SM00355; Znf_C2H2; 5.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 5.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 2282 AA; 246919 MW; E4404B8C9F7B2C CRC64;
```

Query Match 6.6%; Score 346.5; DB 11; Length 2282;
Best Local Similarity 23.0%; Pred. No. 6.1e-12;
Matches 224; Conservative 114; Mismatches 336; Indels 299; Gaps 44;

```
203 RQVLLGSLGQTVGAPASPELPGTGA---SSTKPL--PLFSPKPVQTSKTLASS 255
18 RKRLTKGEALIQTSVSSAP--YPGSGTTAPSESATQELLAQPPSGP-----SQ 64
256 SSSSSSSSGAETPK-QAFHLYHPLGSHHPFSAGVGSHKPTPA-----PSPALPGT 308
65 EKTGOOKPARRPSIEASVHI-----QLP-----QHPLTPAFMSPGKPEHLLEGST 111
309 DOLI---ASPHLAF-----PSTTGLAACQL-GAARGLEATASPELLPKNG 351
112 WQVDPMRPBGSGSFAPGSHPOSQLLPSHASSLPPEELPGLPKVFVPPSOVSLRP-- 168
352 SGELSYGEVNGPDEKPGHRCRCFCAKVFGSDSALQIHLSRHTGERPYKCNVGNFTTR 411
169 -AEENHKKERK-QKR-GKTYCOYCSRPCAKPSVLQKIIRSHGTGERPYCGPGCFEKT 225
412 GNLYVHHRHREKYPHVQNPHPVPEHLDYVITSSGLPYGMS---VPPEKAE----- 461
226 SNL-----YKHKRSHAIRIKAGLAGSSSEMYPPEGLMERIPGEFFEPTEGEST 275
462 -----EAATPGGGERKP-----LVASTALSTESLTILSTAGTA---TAPGL 503
276 DSEETGAAGSPSTDVLPKPNLLSSSLYSSGSHSQERCSLSOSTGPPSLIEDPAPFA 335
504 PAFNKFVL-----MKAVEPNKADENT--PGESEGAISGVASTATLM 546
336 EASSEHPLSHKPEDHTTIQKIALRLSEKKKLIIEQFTLSGSKSTEGSYRSASAQ 395
547 QLSKMLTSLPSWALL-----TNHKSSTGSPF-LPLCAR---ALGASPSSETSKL 590
396 QVSPENTAKSYAEILFGKCGRIGORTSMLASTQPLPLPISSEDKPSLVLPSVPPTQVI 455
591 QQLVEKIDROGAVAVTS-----AASGAPTSAPAPSS----- 623
456 EHITKLITINEAVVDSEIDSVKPRRNSLTRRSSVESPKSSLYRSLSSHGEKTOQDOL 515
624 -----ASSGNOCVYICLVLSCPRALRLHYGOHGERPEPKCKYCGRAFSTRGNRAHF 676
516 LSLQHPSPSTHP---VPLLRSHMPSA-----ACTISTHHTHTRGST--SF 556
677 VGHKASPARAARQNSCPICQKFTNAVTLQOHVNR--HLGQIIPNGGALPBGGAOAEN 733
```

Db 557 DDHVADEVPSPNT--PV-----FTS-----HPRMLKRHAALPLG-----592
QY 734 GSEOSTYVGASGSPPOOQSOQSP--EBELSEEEDEDEEDVD-----ED 779
Db 593 -----EYSSEEPSPSSKDPSTSKPSEDEPERKEDLTKTKKGEKTKANEC 638
QY 780 SLARG-----SESGEKAISVRGDSSEASGAEEVGTAAATAGK--E 822
Db 639 TIGCARYKKRNDYEAKHKYCCSELOITKAHSVGAHEVEKTAEPERPOMMXYKIGATLE 698
QY 823 MDSNEKTTQOSSLPPRRPDSLDOPQPMEOGSSGVLGKEEGKPERSSPASALTPEGE 882
Db 699 LTPLRKRREKSKLEDEEERPAFCPPSEETAHNRPGLSTK---SPAFAKSPASL-EDPR 754
QY 883 ATSYTLVELSLQEAARKPERGESSRKACVCGAFPSQALAEHQKHPRKSPLETGVF 942
Db 755 ASSPPLPSQ-----ELGQNGRRG--EQCPKKTIVYIQTHTSFEKSDPRDP-----798
QY 943 CROGFLEATLKKHMLLHNOVFAPRG-----PQNTAALSLVPGCSPSITS 990
Db 799 -----SGLEEDKPRPAQSPSPRRPHGRSAHSLQRLVYKQNPVQVPELITVEPRDRD 850
QY 991 TGLSPFRKDPPT 1003
Db 851 TEPEPPPEPEKT 863
RESULT 22
Q9NS43 ID Q9NS43 PRELIMINARY; PRT; 751 AA.
AC Q9NS43;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical zinc finger-1-like protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP
RA Kodoiyanni V., Ge Y., Severin J., Krummel G.K., Grable L.,
RA Kviststad E., Gordon L., Brower A., Olsen A.S., Smith L.M.;
RA "Sequence analysis of a 1mb region in 19q13.2 containing a zinc finger
RT gene cluster";
RT Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AC074331; AAF88107.1; -.
DR HSSP; P08045; 1ZNF.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; ZF-C2H2; 18.
DR PRINTS; PRO0048; ZNCFINGER.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; ZnF_C2H2; 18.
DR PROSITE; PSS00805; KRAB; 1.
DR PROSITE; PSS00028; ZINC_FINGER_C2H2_1; 18.
DR PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 17.
DR DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
KW NON_TER
FT SEQUENCE 751 AA; 86207 MW; 96583BAA581C87FC CRC64;
SQ
Query Match 6.58; Score 342; DB 4; Length 751;
Best Local Similarity 21.18; Pred. No. 4e-12;
Matches 187; Conservative 81; Mismatches 335; Indels 282; Gaps 29;
QY 170 VSGSHNIPILIE--LRLVQQRQIHQMOMTEQICROV-----LLG-----209
Db 42 VSGKHQNMKMTLQKALKYLSNQLSCWQIMKWAVASLETRCLQGKSSQLLQDSIOYSEN 101
QY 210 -----SLGQTVGAPASPSSELPGTGTASTKPL-----237

Db 102 ENNINPKGDSSIIYIENOEFPFRMTOHSCGNTY---LSEQIQSRKQIDVKNINQIHEH 158
QY 238 -----PLSPFKPVQTSKTLASSSSSSSSSGAETPKQAFHLYPLGSO--HPFSAGVG 291
Db 159 FMKKSPFHHIKTDTPEKPKGNEYGIIIDGSMOKL-----PLGEKPRPGCGGNG 210
QY 292 RSHKTPAPSPALPGSTDLIASPHLAFPTTGLLAQCLGARGLEATAPGLKPK-- 349
Db 211 FSY-----SPRLPLPHN--VHTGKCFQSQSHLRTHQRIHPGEKLNRCHEGSDCEPKSS 262
QY 350 -----NSGGEISYGEVMGLEKPGGRHKCFKCAKVFESDSALQIHLRSHTEBRPKCV 403
Db 263 FHSYOSNHTGEKSY-----KCDSCGKGFSSSTGLIHYRHTTEBKPKCEE 308
QY 404 CGNRFTTRGNLKVHFRHNRREKYPHVQMNPHVPEHLDVITSSGLPYGMSV-----454
Db 309 CCKCFQSQSNFQCHQVHTTEK-----YKCECGGFGFSGVNLRVHQRVH 354
QY 455 ---PREKAEENATPGGVERKPLVASTLATESTLTLSTAGTAPARGLPAFNKPYL 511
Db 355 RGEKPYKCEE-----YCGKGFSSHNSPLICHRVHTGEKPYKCEACGKFT 377
QY 512 MKAV---EPKNKADENTPPGSEGAISGVAESSTATIMQSLMTSLPSMAL-----560
Db 378 HQRVHTGEKPYKCD-----YCGKGFSSHNSPLICHRVHTGEKPYKCEACGKFT 426
QY 561 ----LTNNFK--STGSPPLPLCARALGASPSFETSKLQQLVEKIDROGAVAVTSAAGAPT 615
Db 427 RNTDLIHFRVHTGEKPYK--CKEKGKFSQASNLQVINDVHTGEKPYKCEGKFSQS 484
QY 616 SAPAPSSASG---PNOVCICLRVLSCPRALRLHYGONGERPFKCVCGRAFSTRGNLR 673
Db 485 SKLQTHQVHTGEKPYKCDYCGKDFSYSNLKLHQVHTGEKPYKCEEGKGFSSWSNLH 544
QY 674 AHPVHGKASAAARQNSCPICQKFTNAVTLOQVHRMHLGGQIPNGTALPEGGAQEN 733
Db 545 AHQRVHTGEKPYK---CEODKSFQALIDFRVHQVHTGEKPYKGCVC---GKGSQSS 597
QY 734 GSEOSTYVGASGSPPOOQSOQSP--EBELSEEEDEDEEDVDDEDLSLAGRSESGEKA 793
Db 598 GLQSHQVHTGEKPY-----KQDVCGKGFGRYSQFI 628
QY 794 ISVRGDSSEASGAEEVGTAAATAGKEMDSNEKTTQOSSLPPRRPDSLDOPQPMEOG 853
Db 629 YHQRGHTGEKPYKCEEG---KGFGRSL-----NLRIHQVHTG 664
QY 854 SSGVLGKEEGKPERSSPASALTPEEATSVTLVELSLQEAARKPERESSSRK--AC 911
Db 665 EKPHI--CEECKG-----AFSLPSNLKVLHGVHTREKLPKC 698
QY 912 EYVCGAAPPQALAEHQKHPRKEGPLTVCFCROGFLEATLKKH 956
Db 699 EECGKGFSSQARLEAHQVHTGEKPYKCDIDCKDBRHHSRLVYH 742
RESULT 23
Q9UEG4 ID Q9UEG4 PRELIMINARY; PRT; 927 AA.
AC Q9UEG4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE KIAA0326 protein (Fragment).
GN KIAA0326.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97349984; PubMed=9705841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,

566 HMDENLFVCSDCGKAFLEAHELEQHR-----VIHERGKTPARRAQ----- 605

```

Db      158 HQRGHTGERPYACADCGKSFADPS---VFRKHRR--THAGLRP-----

```

```
Db      158 HQGHGTERPYACDCGKSFADPS---VFKHRR--THAGLR-----YSCERC GK 203
```

Db 158 HQRGHTGERPYACADCGKSFADPS---VERKHRR--THAGLRP-----YSCERCCK 203

OY	449	PGMSVPRPEAKEEEDATPGGCVKEKPLVAST--TALSTESTITLSTAGTATAPCLPAP	506
Db	204	AVA-ELKDLRNHRSHT-----GERPLCSGCKSFSKSSSLTCHORHIAAKPYKCPAC	257
OY	507	NK-FVLMKAVEPKNADEMRPGSEGSALSGVAESSTATLMQSLKMTSLSPWALLTNPF	565
Db	258	GKGFTOLSTQSHERTHSGKP-----	279
OY	566	KSTGSEPLPLCARALGASPETSXLQDLVEKID-----ROGAVALVTSASGAPT	614
Db	280	-----FLPCRCGR-MFSDPSSFRHGOAHNEGVKPYHCCKCKDFRQ-----PA	321
OY	615	TSAPARSSASSSPNOCVITLVLSCRALYLYHGNGEGRPFCKVCGRASTKGNLA	674
Db	322	DLAMHRRVHTGDRPFKLCLODKTFEVAWMDLKRHALVHNGORFRCCEGCRAFAEASLTK	381
OY	675	HEVGHKASPARAARONSCPRCKKFTNVTLDOHNMHNGO---PRNG-----GTALPBG	726
Db	382	HNKVN-----SGERPHCNACGKSTVSVSSLSLKHERTHNSSEAGVPPRADELVLGLAPV-	436
OY	727	GGAAGENGSEOSTVSGAGSFPPQOQSOQSPREEELSEEEDEDEEDVTDSDSLGRGS	786
Db	437	-GVAGESSAARAGACGLDPPA-----GILGLRP	464
OY	787	ESGGEKA-----ISRGSEELASGAEEVGYAAATACKEMDSNEKTTQOSSLRPP	839
Db	465	ESGGVATMOQVVGMYTEVHEECODAGVREARPLLEGAGE-BADEK-----	512
OY	840	PRDLDOPRPMEOGSSGVLGKGEGRKERSSSPASALTPEGEATSVTLVEELSLQEAR	899
Db	513	PRQFVGR-ECKEFSTTLLRNHRSNRLRPP-----	545
OY	900	KEBGESSRKACEVCGAARPSOALLENOKTHPEKGRPLTCVCSROGFLERATLKKHML	959
Db	546	-----CTCCGKSISDRAGLRKHSRTHSSVPR-YTCRNCRAFLSADLRKNE--	591
OY	960	ANHOVRAPRPHNGRONIALSLVRCSSPS	967
Db	592	RTHVRPMGTRPLRLVALLGMPEEGR	619
RESULT	25		
O960L6			
AC	O960L6	PRELIMINARY; PRT; 736 AA.	
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	LD4623339.		
GN	CG12299.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Prosoiphila.		
OX	NCBI_Taxid=7227;		
RN	111		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY;		
RA	Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,		
RA	Champe M., Chavez C., Dorsett V., Farfan D., Fiske E., George R.,		
RA	Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,		
RA	Nunoo J., Pacleb J., Pargacs V., Park S., Phouanavong S., Wan K.,		
RA	Yu C., Lewis S.E., Rubin G.M., Celniker S.,		
RL	Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AV051994; AAK93418.1; -		
DR	FLYbase: FBgn0032295; CG12299.		
DR	InterPro: IPR000822; Znf.C2H2.		
DR	Pfam: PF00096; zf-C2H2; 10.		
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_10.		
DR	PROSITE: PS0157; ZINC_FINGER_C2H2_2; 10.		
DR	DNA-binding; Zinc-finger.		
SO	SEQUENCE 736 AA; 81670 MW; 0B1C198D547B4C2C CRC64;		

Query Match	6.3%	Score 333.5	DB 5	Length 736
Best Local Similarity	21.0%	Pred. No. 1.2e-11		
Matches 179	Conservative 82	Mismatches 279	Indels 31	Gaps 31
QY 142	SPRLGATPLPPESTRPAPRRPPRRPPRGVSGSLNT-----PLT-ELRLVQDQRIDH	193		
DB 17	TPLLQDTHLLPQAAPNPQQPQQPQQPRLDTEFCMCAAEFVHRLALYQHNHTLHPHPG	76		
QY 194	OMO-----MTEQICRQVLLLSIGQTVGAPASPELPGTGAASPTKPLP	238		
DB 77	NGQOEQSSPDDESEDYKSMIFEPVC-----ELAEGSDSDSGSASSGSDSSSD	125		
QY 239	LFSPIKEVQTSKTLASSSSSSSSSGAETPKQAFPLHMLGSONPES--AGVGRSH--	294		
DB 126	-----DDDDDDDDSSSSSSSSSSSVPTNSMTQGSQESVQPLHGLVAGPQYNEFQ	183		
QY 295	-----KPPAPSPALPGSIDQL-----ASPHLAFPTTGILAAQCLGA	334		
DB 184	LQMTDPRESTSIFWQPVSVYPR-----LQQLLPAPVSPGLQISTP-----IKRR	231		
QY 335	RGLAETASPGLLPKR--NGSG-----ELSYGVNGPRLPKPGGRH-----KCRFCAYV	379		
DB 232	RGRSRNIGAPYMDALNQGCKFOCTHCASFPNA--GDLSKHVRSHTLTKPFQSCIOKT	290		
QY 380	FGSDSALQIHLRSHTEGRRPKCNVCNRPRTTRGNLKVHNRHREKYPH--VQMPHPVPE	437		
DB 291	FTIHGSLNTHIRIHSGEKRPKCELPKATQSSSLWVHMHSVRKHQCYQCKG-----	346		
QY 438	HLDYVTSSGLPYCKMSVPERKAEEAATPGGVERKPLVASTALSTESTLTLSTAGT	497		
DB 347	-----FIWYSSL-----LHMQKHIAPLETF-----	367		
QY 498	ATAPGLAFNKFVLMKAYEPKMKADENTPGRSGSAISGVAESSTATLMOGLMTSLPS	557		
DB 368	-----ICPECEFEKAE-----	379		
QY 558	WALLTNHFKSTGS---FPLPLCARALGASPSSETSKLQOLYEKIDROGAVAVTSAAGAPT	614		
DB 380	-ALDEHMHTQDELVYGCALCREAFRAS-----SELVQHMKNH-----MGEKRF	423		
QY 615	TSAPAPSSASSG-----PNOVCICRLVLSCPRALRYLHGYONGGERPKCV	661		
DB 424	TCSLCDSFTQSGSLNIMIRIHTGKEPFQCKLCDCFTQASSISLVHMKIHAGEKRPYCPRI	483		
QY 662	CGRAFSTRGNRAHFVGHKASPARAQN-----CPLOQKRTNTAVTLOH	707		
DB 484	CGKSYSOQAVYLNKHQIAQOMASASASTSPGLLVAKOPHETLYVCISGLHADATLASH	543		
QY 708	VR-----MHGGOIIPNGSTALPEGGAQENES-----EQ	737		
DB 544	VHGHQALLDTMKSGMNTAPRGALIPDVGCSAEBOQAYVERVQCVLODMNQOQHQQOQ	603		
QY 738	STVSGAGSPFOQGSQO-----PSPPEELSEEBEED	768		
DB 604	QRPQQQQQHPPQQQQQQHLLQCPHMQPLPQOPKPLPAMDSTGEDEEERPDAAEPPDEEED	663		
QY 769	EEEE-----EDVYDEDSLAGSGESG--EKRAISVRG-----SERASG--	805		
DB 664	EPERPAEVKTEVLAAEDALTPNGYPIILGLOEQIILLDSQWYEDFQDMVQCGQEEVFEDF	723		
QY 806	--AEEEVGTVA 814			
DB 724	VVNEEEVYTTDA 734			
RESULT 26				
ID 015090	PRELIMINARY:	PRT: 1300 AA.		
AC 015090				
DT 01-JAN-1998	(TREMblrel. 05, Created)			
DT 01-JAN-1998	(TREMblrel. 05, Last sequence update)			
DT 01-JUN-2002	(TREMblrel. 21, Last annotation update)			
DE KIAA0390	Protein			

RX MEDLINE=98087435; PubMed=9427563;
RA Miyake J.H., Szeto D.P., Stumph W.E.;
RT "Analysis of the structure and expression of the chicken gene encoding
a homolog of the human RREB-1 transcription factor";
RL Gene 202;177-186(1997).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AF013754; AAB9584.1; -.
DR HSSP; P08045; 12NF.
DR InterPro; IPR000822; znf.C2H2.
DR Pfam; PF00096; zf-C2H2; 15.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00355; znf_C2H2; 15.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 13.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SO SEQUENCE 1615 AA; 178279 MW; 5C2F5F0C0FA16C79 CRC64;
Query Match 6.1%; Score 322; DB 13; Length 1615;
Best Local Similarity 18.9%; Pred. No. 1.5e-10;
Matches 268; Conservative 153; Mismatches 457; Indels 538; Gaps 57;
QY 4 ESEBSRLGVPRPPELG---GDASEED--HPOYCAK----- 36
DB 160 EMDREER--TPRAKVEDGQYGEGRKEDDAYHCPVCFKCYALASHMETHPNLSLR 217
QY 37 ---CCAQFTDPFEFLAH---QNACTDPYVAVITGGENPN-----NSSASSE 78
DB 218 CDICCTFTFTHRLRLRHNAVHKLPRDPTGKPF--QNNPSIPAGFHLDGTFDFSCRF 275
QY 79 PRPEG-----HNNQVMDTEHSNPPDGSSV 104
DB 216 PRISQVWCETNLRCISEFHRIETCNCAPMLLALKLHTEHVMDQGRDKHKLSTSL 335
QY 105 PRDPTMGPRRGESSHFLVAATG-----TAA 132
DB 336 PSE---NPOKA-----FMASLGLOYTKDKPKVKEODTODEVOEMRLALKSNLPQE 385
QY 133 GGGGGLILASPKGLAPLPPESTRAPPPPPPPGVSGHNLPILELHVLQORQI 192
DB 386 PSTGTLISLPLEAATMGPFVSV-LPTKE-----NIKLL-SLPFGKGF 429
QY 193 HQMQMTEQI-----CROVLLGSLGQTVGAPASPELRGTASTTKPLPL 239
DB 430 IQPDSSIVAPKISNESAIELADIQILKMAS-----SAPQIIS-LPPLSKAPSV-PVQSI 482
QY 240 FS---PI--KPVQTSKTLASSSSSSSSSSSGAETPKQAFHLVHPLGSSQHFSGVGRSH 294
DB 483 FKHMPPLKRPPLVTPRTVAVTS-----TPP-----PLISAQOASPGCI---- 520
QY 295 KPTPASPALPGSTDQLI-----ASPHLAFSTTGLLAOCIGAAR-----G 336
DB 521 -----SPSLRPPPLRLIKKNVSESSNSHLSOPGAKSSPSQLDLPKVPRLQIHEMKTQ 574
QY 337 LEATASPGLLKPKNGSELGYEWMGLEK--PGGRHK-----CRCAKVFSGD 383
DB 575 LEQDSITTEALLPLIMEAKIKQETEGEDLKAITAGAAANKKAPTRKVLVPCRFCDQVAFS 634
QY 384 SAQIHLRSHGTGERPYKCNCV-----GNR-FTTR-----GLKLVHFN 419
DB 635 GVALRAHIRSLHGISPYOCNICDYTAADRAALIRHLRTHSGERYICKICHYPTVKANCE 694
QY 420 RHREKPRHVQMNPRVPEHLDYVTSSGLPYGMSVPERKA-----EEEAAT 465
DB 695 RHLRK-KHLVTRKDIEKNIEYVSNAAEMVDAFCSPDVCKLGEDLKHVRLAIRIMRT 753
QY 466 PGGGVERRP-----LVASTAATSLT 489
DB 754 HSGCQKKKPECKECSGTAFSAKRNCIHHILKOHLYQEREIENYIIVDSCSAQSHDAP 813
QY 490 LLSSTAGTAAAGLPAAF---NKFVL-----MKA 514
DB 814 LLEDSTYIMDCRPITPFLERPNQFSIGTSSHVPIKLEPMGNFPMDFDEPLDPSQKSKNLSA 873

QY 515 VE-----PKNKADENTPGSE----- 530
DB 874 VQYKQENLVSPLSFYDCSMEPIDLSIPKYLKRNDDIPGEARNQELASSVITDNVNCQ 933
QY 531 ---GSAISGVAAESSAT-----LMQ 547
DB 934 QCPILGPGANSNSEKNRAVGHQDPLKGSLLHVPITISPLLGNALLRPLRPKPPQPLP 993
QY 548 LSKIMTSLPFWALLTNFKSGSPFLPCARLGAASPSFTSKLOQLVEKIDROGAVAVTS 607
DB 994 KPPVTKELPPLASTIAQIITSSVSSAPALLKTEAADAASPRAASSSTGCDKGNAKAKMTIVT 1053
QY 608 AA---SGAPT--TSA--PAVSSSASGPNOCVTCRLVLSCPRALRLHYONGGERPPCK 660
DB 1054 AIGRDSNLPEDLIQACDPEPPIADTG-----LTGKRGKKGSTKN-KPK 1096
QY 661 VCGRA-----FSTRGNLRHNFVGHKASPARA-----QN----- 689
DB 1097 LSSGVLDLESSGEFASIKKMLATTDITNKSPFLQSTDNFKESGQNGTSEDEKETPEDKL 1156
QY 690 -----SCPICKKFTNAVTLQOHVNRHILGGQIFNCGTALPPEGG----- 728
DB 1157 RGRNNTSDCPQKITTCTCYCPVFSWASSLQRHMLHTDSQ---ADTEAPATGGEVLDLTS 1213
QY 729 AAOENGSEOSTVSGAGSPPOQSOQSPD--EBELSEEEEBEEDDEED----- 774
DB 1214 CEKEQPEVESELPESECSPOEQKADSPPAEDAEKKADEYEGREDPSVSNKSLDLNFA 1273
QY 775 -----VTDEDSLAKGSESGEKAISVRGDSSEKSGA--EEEVGTVAATAAGKEMDS 825
DB 1274 SKLMEFKLASDSQAGSSQTERKHACDVCCTKFRGALSRKKAHIREDRDERSED 1333
QY 826 NEKTQOSSLPRPPPSLDQPO--PME--OGSSGVLGKKEGKPERSSPASALTPEGEA 883
DB 1334 ESKSIDDAQAPSNQDSGLEBESPMDLKVYSPPLDCEATGKNESES-----ISGEG 1388
QY 884 TSVTLVELSLQEMRKREGESS-----RKACEVCGAAPPQAALEHOKTJPKESPLF 938
DB 1389 TERKSTESDDKIPKTDKASTAKADRKRYCTVQCNKRFSIQLDILRHMRSHTERP-Y 1447
QY 939 TCYFCRQGFLEATLKKHMLAHNOVPRAPRGPON 974
DB 1448 KCQTCERTFTLHSLVRRQRI-HQKVKNTRNHGKES 1482
RESULT 31
Q96I71 PRELIMINARY; PRT; 734 AA.
AC Q96I71;
DT 01-DEC-2001 (TREMBLrel, 19, Created)
DT 01-DEC-2001 (TREMBLrel, 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)
DE Unknown (protein for MGC:12654).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC007777; AAH07777.1; -.
DR InterPro; IPR003309; Treg_SCAN.
DR InterPro; IPR000822; znf_C2H2.
DR Pfam; PF02023; SCAN; 1.
DR Pfam; PF00096; zf-C2H2; 13.
DR ProDom; PD000003; znf_C2H2; 6.
DR PROSITE; PS50804; SCAN_BOX; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_13.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 13.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 734 AA; 82055 MW; B804766A0B008048 CRC64;

Query Match 6.1%; Score 321; DB 4; Length 734;
Best Local Similarity 20.7%; Pred. No. 6.6e-11;
Matches 196; Conservative 74; Mismatches 309; Indels 368; Gaps 37;
80 RPEGHNPQVMDT-----EHSNPPDSSGVPTDPTGPERGESSGHFLVAATGAC 133
b 74 RPEVRSKEQMLLELVLEQFLGALPEELIARVQ-----GQRFSGPEEAAALVDGLRREP 127
y 134 GG-----GGILASPKIGAT---PLPEESTAPPPPPPPPPPPGVSGLHNTPLILE 182
b 128 GPRRWVVOVQGVSEKMEPSSFOPLPETEPPPEGPKTP----- 171
y 183 ELRLVQOROI-----ROMQTEQICROVLLIGSGQVGAAPASELPETGTAATSTKPL 237
b 172 --RTMQESPLGLQVKESSEYTED-----SDPLESGPLAATQESV 208
y 238 PLESPKPVOTSKTLASSSSSSSSSGAETPKQAFHLVHPLGSOHPFSAGVGRSHKPT 297
b 209 PTLPEEAORCGTVL--DQIFPHSKTGPESGSMR-----EHPRALMHE--EAGGTF----- 255
y 298 PAPSPAL-----PGSTQQLLSPHLAFPTSTGL--LAQCCLGARGLDEATASGLLK 347
b 256 --SPGRLAQGSISAGSGS-----VSPHLHVPMWDLGMAGLSQIQSPSR--EGGFALALL 307
y 348 PKNGSGELSYGEV-----MGPL-----EKPGGRH-----KCRFCATVFGS 382
b 308 PSDLSQEDPTDEDPORGVPALITTRMRSPRGSRGRPSGTGGVVRGRCQVCGKVFESQ 367
y 383 DSALOIHLRSHGERRYKCVCCGNRFTTGNLKVHFNHREKPYHVQNNPHVPEHLDIY 442
b 368 RSNLLRHQHIHGERFVSECCRSFSSHLRLHQLHTTERP-----FV 413
y 443 ITSSGLPYGMSVPEPEKAEEBAATFPGGVVERKPLVASTALSTESTLTLSTAGTAPG 502
b 414 CGDCQGF---VRSALIEHRVHTG---EOPFRCACGGQSFROSNLLOHORIHDPPG 467
y 503 LPAFNKFLVLMKAVERPKKADENTPRGSESAISGVAESSTATLMOLSKLMTSLPWSALL 562
b 468 -----PGAKPPAPGAPE----- 480
y 563 NHFKSTGSPFLPCARALGASPSSETSKLOLVEKIDROGAVAVTSAASGAPTTSAAPSS 622
b 481 ---PPGPFCSECR-----ESFARRAVLLEHQAVH-----TG 509
y 623 SASGPNQCVICLRYLSCPRALRLHYGOHGERPFCKVCVGRAFSTRGNLRAHFVGHKAS 682
b 510 DKSFG---CVEGGERGRSVLLQHRVHSGERPFACACGCSFRORSNLTQHRRIHTGE 566
y 683 -PAPRAQNSCPTCQKKFTYNAVTLQOHVBMHLGQIPNGGTALPEGGAAQENGSEOSTVS 741
b 567 RPFV-----CAEGKAFRORPTLQHLRVHTGK---PFCAPCEG----- 603
y 742 GAGSPQQOQSPRPEELISEEEDDEEDEDVTDSDSLAGRSESQGEKALISVAGDSE 801
b 604 -----QRFQRIKLTLRHQRTH-----GK----- 623
y 802 EASGAEEVGTAAATAAGKEMDSNEKTQOSSLRPPRPDSDLOPOMEQSSGVLGK 861
b 624 -----QRFQRIKLTLRHQRTH-----GK----- 623
y 862 EEGGPERKSSPASALTPEGEATSVTLVEELSLQEMARKEPGESSSKKACVCGGAPSO 921
b 633 -----GFTQVSRILT--EHORIHGGERPF--ACRPGCGQFRQH 665
y 922 AALEHOKTHPKEGRLFTGVCFCROGFLERATLKKHMLLAHQVQRPFA 968
b 666 ANLTQHRRIHTGERP--YACRPGCGKAFRORPTLQH--LTHHREKPF 710

AC Q924S6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Zinc finger protein 219.
GN 2010302A17RIK OR ZNF219.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Sakai T., Maeda H., Hayashizaki Y.;
RT "Identification and analysis of the alternative promoter regions of
RL the human and mouse ZNF219 gene.";
RN Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=20499374; PubMed=11042159;
RA Carrinci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper selected cDNAs to
RL prepare full-length cDNA libraries for high-rate new gene discovery.";
DR EMBL; AB063578; BAB61057.1; -;
DR MGD; MGI:1917140; 2010302A17RIK.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_5.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 726 AA; 77857 MW; 82A51328FE6B11E CRC64;
Query Match 6.0%; Score 316.5; DB 11; Length 726;
Best Local Similarity 21.8%; Pred. No. 1.2e-10;
Matches 163; Conservative 83; Mismatches 280; Indels 223; Gaps 27;
y 304 LRGSTQQLLASPHL--AFPSNTGLLAQCCLGARGLDEATASGLLKPKNGSGELSYGEVM 361
b 1 MEGSRRIILVG-HLESPFPAPDGLDLQRTSNGFVSGTPTGG-----SFGMGQVGSSEF- 54
y 362 GPLEKPGGRH-KCRFCATVFGSGDSALOILRSHGERPYKCVCCGNRFTTGNLKVHFNH 420
b 55 ---RAGERFRPCVCGKRRRNSIILALHRAHPGAQAFQPCGCHRAORALLRSHLRT 110
y 421 HREKTYHVQNNPHVPEHLDIYVITSSGLPYGMSVPEPEKAEEBAATFPGGVVERKPLVAST 480
b 111 HQ-----PERRSPFAARLLLELEERALLREAR 137
y 481 ALSATESLTLLSTAGTATA--PGLPANRKFVLMKAVERPKKADENTPPGE----- 530
b 138 LGRARSSGGMOSSPAEGLARPOVPSSAIFRC-----PFCKKFTSARERLHILHRP 192
y 531 -GSAISGVAESSTATLMOLSKLMTSLPWSALLTNHFKSTGSPFLPCARALGASP----- 584
b 193 WKCSLCSFGSSQDEELIHN-----LTAN-----GASERLAAATSTDEPPPPQ 237
y 585 SETSKLOQLVE-----KIDROGANAVTSAASGAPTTSAAPSSASSGNGQCVICLRYLS 639
b 238 EPRSALEPEDEPPRPDR-----ANRAPP--PAPPEPPAPPE----- 276
y 640 CPRALRLHYGOHGERPFCKVCVGRAFSTRGNLRAHFVGHKASPARAQNOSCPTCQKKFT 699
b 277 -----FRQVCGQSTQSWFLGHHMKHKAS-----FDHACRPGVGRCKR 315
y 700 NAVTLQOHVBMHLGQIPNGGTALPEGGAAQENGSEOSTVSAGSPFOQSQSP----- 754
b 316 EPWFLLKNHKVHTSKLGP-----LRAPGPSAPARAPQPPDLSTLL 355
y 755 -----SPEELISEEEDDEEDEDVTDSDSLA-----GRGSSGGEKAL----- 794

QY	197	-----MTEDICRQVLLLG-SLG-QTGAAPASPEL-----PGCTASS	232
Db	653	TGSTFTVAEVTQSQTIVSGSGLGQSTNSPSPSLSPSTSGCMSTLTSEPPSSIQSSG	712
QY	233	TKPLPLPSPKIPVQTSKTLASSSSSSSSGAETPKQAFPHLYHPLGSOHPFSAGVGR	292
Db	713	AQSTLTFTSP-NPQSTSLSEBSTSGATTSSSGSAGTTM-----SPQSSSVGS	760
QY	293	SHKPPAPSPALPGSTDOLIASPHLAFSTGLLAOCGLAARGLEATAPG--LLKPKN	350
Db	761	SQGST--SPAASTSGEMTSQGSTQTPGSSVSTSAALTSTQOVSIVTNSPGSTVTPRST	817
QY	351	GSGLSTYGE--VMGLEKPGGHHKRCFKAVYGSLSALQIHLSHNGEPRYKCNVGNPF	408
Db	818	VSGSTSSGSTVTVGTE-----ASTGSSVYA-----	843
QY	409	TTRGNLKFHFHRRKRYPHVQNPMPVPEHLVLVTYSGLPGVMSVPEKAEATPGG	468
Db	844	-----SSSPAPSTSNPNPNTSGSSSMITQSYPPQSTP-----VSSITTPP	887
QY	469	GVERKPLVASTALATESLTLSTAGTATAPGLPAFNKPYLMKAVEPKNADENTPPG	528
Db	888	GSPGTTL--TSTSPSGQSTTGISNQG--STSPGISTGS-----EEMTSQSTQTPG	935
QY	529	SEGSAL---SGVAES-----STATMQLSKMLSLPFWALLTNHFKSTGSPPLPCARLG	581
Db	936	SGSTFTVTPSTVSDSTSSGSTVTVGTEGSSSPIPSTONTNPSTSGS-----	984
QY	582	ASPSTSKDQOLEKIDRGAVAVNTAAGAP---TTAPAPSSASSGPNOCVILRY	637
Db	965	SKSTQTPQSSQSTSPVESSTSGATSS--SGSPQTTLTSLSPSPSPSTIGSSQGSTSPV	1042
QY	638	LSCPALRLHYQHOGGERPFCVKVCGRAFS---TRGNLRAHFV-----GHKASP	683
Db	1043	STISQSTETPSTGSTVTKPFTVSGSASSGATATMGSTFASSTSGSSSTSPMPQSSTSP	1102
QY	684	AARAONSCPTCKKFTNNAVTLQOHVAMHLG-----QIP--NG	719
Db	1103	STSGATSSPGSSGTTLTSLSPSPQSTTIGSSQGSTSPVVTTSQDMTSQSTQIPGSGT	1162
QY	720	GTALEPGGAQENGSEOSTVSGAGSFPOOQSOQSPPEELSEEEDEEDEDVTD	779
Db	1163	STVTPQSTGSGSTSGELTSGQSTTP--RSLSTSPALSTSQOASVTNPSQSTVTPQS	1221
QY	780	SLAGRSSESGEKAISVRDSEBASAEBEVGTVAANAATAGKEMDSNEKTTQOSSLPPP	839
Db	1222	TV--RQSTSSG-----STVTGSTEGSSTGSSSATLSLS--SSP	1257
QY	840	PPDSLDDQPMEDGSSGVLGKKEGKPERSSSPASALPREGATSVTLVEELSLQDAMR	899
Db	1258	VPTSGSPMPSTSGSS-----TPPNPQSTSPVVS-TTGGMTSHGSTQTPSTIGSTV	1310
QY	900	KEP-----GSSSRKACEV-----CGQAF--PSQALEHOKTHPEKGPLFTCVPCQRQ	945
Db	1311	TOPSTVYSGNSGQSTVTTIGSSSEASTSGSSKRTKTPSSLS-----PVTSSPISTTPASS	1364
QY	946	GFLERATLKHHMLAHQVQFAPHGPQNTIALSLVPGCSPSTST-----GLSPFR	998
Db	1365	--TSGSTIDSVSVSTSLAPLSSLP-----STVPSSQSFSTSEGGSSKASSSPVS	1416
QY	999	K--DDPTIP 1005	
Db	1417	QNSSTPTNP 1425	
RESULT 34			
ID	061360	PRELIMINARY; PRT: 962 AA.	
AC	061360;		
DT	01-AUG-1998 (Tremblrel, 07, Created)		
DT	01-AUG-1998 (Tremblrel, 07, last sequence update)		
DT	01-JUN-2002 (Tremblrel, 21, last annotation update)		
DE	CROL alpha		

QY	115	----	RGESSGHFLVAATGTAAAGGGGLLASPKYGLATPRPESTRD-----	PPRRP	164	
Db	214	TVQVQGIKE-----	-----VLSEKME-----	PSSTQPRPHIKIQTDPPGAE	248	
QY	165	PRRPGVGSCHNLPIRLLEELRYLQDQRIHOMOMTEQICROVLLLSL--	CGTVGARPAS	222		
Db	249	TPRGAMQELRLPM--KEETELLGNPRL-----	LESGRPLAQRDAATLLPK	294		
QY	223	ELRPGTGA-----	SSTRPLRLPFS-----	PIKPVQTSKLTAASSSSSSSSSGAETPKQA	271	
Db	295	ETQGGHTLDTQTSPHSDTEBDVPRWSDRPRLMHEEVGIGISPGFTLQMDSTAE--	PDMM	353		
QY	272	FRHLXPR-----	IGS-----	QHRPSAGVGRSHKFTPARSPALP-----	GSTDOLIASPHLA	318
Db	354	SPHLHPWMDLMAISLSGOISPTTEGTF-----	ANALVLPSPDGEORPACEDP---	402		
QY	319	PRSTGTLAAGCGAARGLEATASPGLKPKNSGSGELSYGVMGRLEKPRGGHKKRFGAK	378			
Db	403	CPSPVGRPALVATRMHARSGSQS-----	HRSGTSGAMR-----	GGR--	CDVCSK	443
QY	379	VFGSDSALQIHLRSHTGEREYKCNVCSNREFTTGNLKVHFNHREKURHYVOMNPRVPEH	438			
Db	444	VFSQSRNLRLHQHHTGERPFVCGEGRSFRSSHLHLNQLHTHEERP-----	491			
QY	439	LDVYTTSSGLPRGMSVPRPEAAEEBAPRGCGVEKEKPLVASTALSATESLTLTSSAGTA	498			
Db	492	--FVCRDCQGF--	VRASLEHHRVHTG--	BQPRCTGCGGSEFRKRSNLLQHORING	543	
QY	499	TAPGLPRAFKFVYLKAVEPRKNKADENTPRGSEGSAISGVAESSTATLMOLSKLMTSLPSM	558			
Db	544	DEPG--PA--	PSILPLAGVPE-----	PPGP-----	PPCSGCCSEFFRR	577
QY	559	ALLTNHFKSTGSEFPLPLCARALGASPESETSKLOOLYEKIDROGAVAVTSASGAPTTSAP	618			
Db	578	AVLLER-----	-----OAVHTGDK-----	591		
QY	619	AVSSASGSGNOCVLCRLVLSCPRALRLHYGQHGGERPFCKVCGRAFTSGNLRANFVG	678			
Db	592	-----SFG-----	CVEGGERGRNSVLLQHRVYHSGERPFACACGQSFQRKSNLTQHQNI	642		
QY	679	HKASPARAONSCPIQCKKFTNAVTLQOHVBMHLGGQIPNGTALREGGAOENSGEOS	738			
Db	643	HNGERAF-----	ACAEGCKTFRQRTATLLOHLCVHNGEK--	PFACRPEG-----	683	
QY	739	TVSGAGSFPQOQSQPSPREELSEEEDEEEDVDVDEDSLAGRSGSGEKAISVG	798			
Db	684	-----	-----QRFSPQRKLKLTNRQHTT-----	GEKPYCC--	707	
QY	799	DSEBAGAEVEVGTVAANAATAGKEMDSNEKTQOSSLPRPPRPSIDQRPMBGQSSGYL	858			
Db	708	-----	-----	707		
QY	859	GCKEKGCKPRSSSPASALPRPEGATSVTLVEEUSLODAMRKREPESSSRKACEVCGQAF	918			
Db	708	-----	GECDLGFQTVQVSRLT-----	EHQRIHTGERPF--	ACRPEGQSPF	742
QY	919	PSQALAEHQTKHKEGRPLFTVCFCRQGFELRATLKKIMLLAHNOVOPRA	968			
Db	743	ROHANLTQHRRIHTGERP--YACAECKGAKFRORPITLQH--	LRTNHKEKPPA	790		
RESULT 36						
046205						
AC	046205	PRELIMINARY;	PRT;	1920	AA.	
DT	01-JUN-1998	(TREMBLrel. 06, Created)				
DT	01-JUN-1998	(TREMBLrel. 06, Last sequence update)				
DT	01-MAR-2002	(TREMBLrel. 20, Last annotation update)				
DE	zinc-finger nuclear protein HINDSIGHT.					
OS	PEB OR HINDSIGHT OR EG:6641.1 OR CG12212.					
OC	Drosophila melanogaster (Fruit fly).					
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;					

OC pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=97330681; PubMed=9187140;
 RA yip M.L.R., Lamka M.L., Lipschitz H.D.;
 RT "Control of germ-band retraction in Drosophila by the zinc-finger
 RT protein HINDSIGHT."
 RL Development 124:2129-2141(1997).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; U86010; AAB95640.1; -.
 DR HSSP; P07248; 2ADR.
 DR FlyBase; FBgn0003053; znf.
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 12.
 DR PRINTS; PR00048; ZINCINGER.
 DR SMART; SM00355; Znf_C2H2; 13.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 10.
 KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
 KW SEQUENCE 1920 AA; 208973 MW; BA51E67584916AA2 CRC64;
 Query Match 5.98; Score 313; DB 5; Length 1920;
 Best Local Similarity 20.0%; Pred. No. 6e-10;
 Matches 249; Conservative 137; Mismatches 438; Indels 424; Gaps
 56;
 OY 16 GEPAEIIGDASDEHPQVCAKCAQFPTPTPEFLAHMACSTDPEVMYIIGQENPNNS- 74
 DB 382 GNTLSGAESEYD-----ASCS---TVSSGSHSRSSSS-----LNNNNSH 423
 OY 75 -----ASSEPRPEGHNNPQVMDTEHSPNPDSSGSVPDPTWGPERRGESSG 121
 DB 424 KANNNLKDELEVESTEDQDTEENKQRILKTTINNITIEEQEDMD-----DEADDAV 478
 OY 122 HFLVATGTAAAGGGGGLILASPKIGATPLPEPSTAPPPPPPPPP-----PGVSG--- 173
 DB 479 AMLTSPDVAI-----LLAGASASGA-----ASRPTSPSPASALLSCPACGASDFE 527
 OY 174 -----HLNIPLELERLV-----QORIHOMQTEQICR-QVLLIGSLGTG 216
 DB 528 TLPALCVHLDAMSHDITPAKCDCEVIYFATHROLQH-----CRLPNALAGDPLRLG 580
 OY 217 APASP--SELPGTGTASTKPLPLPLSPIKPVQSTKLASSSSSSSSSSGAETPKQAEFH 274
 DB 581 ASSPPLHNEPEDEHGDDEDL-----EQKRLASQSD----- 614
 OY 275 LHPPLSGQHPFSAGGVGRSHKPT-----ASPALPGSTDOLIA 313
 DB 615 FFLQLYKHTTANCGALSHPPSPPIKHEBADIKDLADIOSILNMTSSSSSPLRFNEQSVN 674
 OY 314 SPLHAFSTGTGLAAQCLGAARGLFATASPL-----LTKP 349
 DB 675 TPMSQVSLDS-----RDQEEADADATSEFRRMKLRGEFFPKCLTAVFPNLRL 724
 OY 350 NGSGELSYGEYMWLEKEDGRHKRCFCAKVEGSDSALQIHLRSHGTGRPYKCNVCGNRF 409
 DB 725 KGNHRRVHIGAV-----GAPGPRCNMCPYAVCDKALVIRHMRTHNGDRPYCAVCNVAFT 779
 OY 410 TRGNLKYHF-HRH-----REK-----YPHQGMNHP 434
 DB 780 TKANCERHLNRHNGTSTREEYKRAIVYHRAEDAGCEDSKSRLEGDLADTSRSTISPTPP 839
 OY 435 VP-----EHLDYVITSSGLPYGMSVPERKAE-----EEAAPPGGVERKPL 475
 DB 840 PPNVESKSQLKHM--LIGENHLAPVNOQPRPKIQVAVSLDGLVKKPSAPAPAQ00000000 897
 OY 476 VASTTALSAVESLTLLSTAGTATAPGLPAFNKFVLMKAVEPKKADENTPPGSEGSAIS 535
 DB 898 EKSGSALDFSDMDVLDLSKRP--TGASILTPAVVTRTPPAVAAPV-----TPGGVGTPLDA 950
 OY 536 GVAVESSTATLMO-----LSKLTMTSLPSPALLLTNHEKS--TGSPF-LPLCARALGAS 583

DB 402 EHLNHNMSHNET-----PRCELCGKSFSSKEH-----431
OY 534 ISGVAESSTATLQLOSLKMTSLPSMALLTNHF-K-STGSFP--LPLCARALGASPSKTL 590
DB 432 -----FTNHLMTHTGETPHRCDCSKTF-----TRK 457
OY 591 QQLVEKIDRGAVAVTASASGAFTTSAPAPSSSSAGSPNOCVICLRVLSCPRALRLHYQ 650
DB 458 EHLNHN-V-RQ-----HTGESPHRCSYCKKTTTRKEHLVNHMQ 494
OY 651 HGGEPFCKVCVGRAFSTRGNLRAHFVGHKASPARARONSCPICOKFTTNAVTLQOHVM 710
DB 495 HTGEFPCFTCYCTKATRD-----HWNHNVRQHTGESHPKCTCYCTKTRKHLTNHVL 550
OY 711 HLGGOIPNGSTALPEGGAANGSEOSTVSGASFPQOQSOPSEBELSEEEDEE 770
DB 551 HTG-----DSPHKCEY 561
OY 771 EEDVDTDSDSLANGRSESGEKAISVRGDSBEASGAEEVGTVAATAAGKEMDSNEKT 830
DB 562 COKTTRKEHL-----NNHMR 577
OY 831 QOSSLPPPPPSLDOPQPMEGCGSVLGCKEGCKPERSSSPASALPPECATSVLYE 890
DB 578 QHSS-DNPHCVCVNCNPFTRKEHLNHNMSRCHTDPR-----FTCECTGCKSPPIKG 627
OY 891 ELISLOEAMKEPGESSSKACEVCGAAPPSCAALEEHOKTHPEKCPPLFTCYRCGFLER 950
DB 628 NLLFQRSHTKGEMERPFACCEKCPKFNCTKGHLVSHMRSHSGEP-NACTLCSKAFAVER 686
OY 951 ATLKRHHMLAH-----HOVOQFPA 968
DB 687 GNLKRMKNNHPDAMPPPPVHHPQIPAGVLTQVKQEVKPIITHHSATTTMTTIOQT 746
OY 969 PHGPONIALSLVPGCSPSITSGLS 994
DB 747 AGAGGAGAVOLTRPLVLTSTLIS 772
RESULT 38
Q9WAJ1 ID Q9WAJ1 PRELIMINARY; PRT; 1893 AA.
AC Q9WAJ1
DT 01-MAY-2000 (Tremblrel, 13, Created)
DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel, 20, Last annotation update)
DE PEB protein (GH10905P)
GN PEB OR EG:66A1.1 OR CG:2212.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
CX NCBI_TaxId=7227;
OY [1]
RN SEQUENCE FROM N.A.
RP STRAIN-BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Aamathides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazey R.G., Chame M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B.G., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazozo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Appayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Fritse E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacle J., Pargass V., Park S., Phouaneavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCRT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003431; AAF45960.1;
DR EMBL: AY058335; AAL13564.1;
DR HSSP: P07248; ZADR.
DR Flybase: FBgn0003053; peb.
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF00096; znf_C2H2_12.
DR SMART: SM00355; znf_C2H2_13.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 11.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_10.
KW DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 1893 AA; 205673 MW; 1B223EE96468A754 CRC64;
Query Match 5.8%; Score 308.5; DB 5; Length 1893;
Best Local Similarity 20.1%; Pred. No. 1,1e-09;
Matches 245; Conservative 132; Mismatches 428; Indels 413; Gaps 55;
OY 53 ACSTDPPVNVIIIGOENPNSSASSEPRREGHN-NFOVMDTEHSNPPDSSSVFTDPTWG 111
DB 397 SCSTD-----VSSGHSRSHSSSLNNNNNSHKANNLKDLE-----ELFVSTEDOOTEN 446
OY 112 PERGESSGFLV-----AATGTAAGGGGGLIILASPLGATPLRPES 154
DB 447 KORRLKTTINNNTIESEQOEDMDDEADADADVAMLTSTPDVATLLGASASGA-----A 500
OY 155 TPAPPPPPPP-----PGVSG-----HLNIPLLIEELRVL-----QQR 190
DB 501 SRBTPPSPASPALLISCPAGASDFETLPALCVHMDAMHSODIPACORDEDEVIFATRHQL 560
OY 191 QIHOMOMTEOICR-OVLLIGSLGTGVAPASP-SELPETGTAASSTKPLLPLESPKPVQ 247
DB 561 QSH-----CCRLPNALAGLPLPILGASSPLHNEPEDEEAGDDBDL-----E 603
OY 248 TSKTLASSSSSSSSSSSGAETPKQAFPHLHPGLGSCQHPFSAGVGRSKPP-----298
DB 604 QKERLASQSD-----FFHQLYLRKHTANGCAISHPSPKIKPEPADTK 647
OY 299 -----APSPALPGSTDOLIASPILAFPSPTGGLAAOCIGARGLAATASPL- 345
DB 648 DLADIQSLNMTSSSSSFLRNFQGVNTNPSOYSLDG-----ROEEBAQDAFT 697
OY 346 -----LKPNGSGELSYGEVMGLEKPGGRHKCRFCAYFGS 382

698 SEPRRMKLRGEPPCKLCTAVFPLRLALKGNRVHVGAV-----GPAGPRFCMMCPYAVCD 752
383 DSAIQIHLRSHTGERPYKCNVCGNRFTRGNLKVHF-HRH-----REK----- 424
753 KAALVHRMRTNHGDRPEYECAVCNVAFPTTKANCERHLNRNHGRTSREVKRAIYHRAEDA 812
425 -----YPHVQNNPHRPV-----EHLDVYITSSGLPYGMSVPEKA 459
813 GCEDSKSRKLGEDLADTSFRTSISPTPPRPVNESKSQLKHM--LLGSHNLAPVNOQRPPLKI 870
460 E-----EEAATPGGVERKPLVASTTALSTATSLTLLSTAGTATADGLPAF 508
871 QVKSIDQIVDKKPSAPAPQ00000000EKSGLALDFSMVDLDSKRP-TGCASTLPAVTR 929
509 FVLMKAVEPKNADEPTPPSGESAIISGVASSATLMO-----LSKIMTSLPSW 558
930 TTPPAVAVV-----TPGGVGTPLDLAALIEQ00LLA00QLFGAGGEYMOQLFRSL-- 980
559 ALLTNHFRS-TGSFP-LPLCARALGASPSETSKLOLVEKIDROGAVAVTSAASGAPPTS 616
981 -----MFGSGTSGFPFPFMA-----PPRPQANPEKPRMVSPRRINIMPV-GVGVGVPV-- 1029
617 APAPSSASGSCNOCYICLRVLCSPRALRLHYG0HGERPFCKVCYCGRAFSTRGNLPAHF 676
1030 -----PPGGVVKVYIKNGVL-MPKOKORRYRT--ERPFACEH0SARFTLRSNMERHV 1078
677 -----VGH-----KASPARA----- 687
1079 KOHROFAQORRSGIHNVARGRASNVAATAAATAATVMAGPGSSGFGSNHHGH 1138
688 -----ONSCPICOKKFTNAVTLQ0HVRMH-----LGQILPNGTAL-----PE 725
1139 GHSGHSGHGHATISEQ--VKCALILAQ0LKAHKNTDILQ0ALAHGSSSVAGNPLHFQYPL 1196
726 GGGA0QENGSBQSTVSGASFPQ00SOOPS-----PBEELSEEBEEDBEE 771
1197 TNPSPMHNSSO-----GNC0ATAMDDEPKLIIIDENENHDEVEDVDDPDEDEDEE 1252
772 EEDVTDENSL-----AGKSES0GE--KAISVRGDSSEASGAEFVGAATAAGKEM 823
1253 MDEPDEDELIIDEQFAKKEAE0E0ELPKPLEJ0LGTREAAQMAETI--LEQAIKAKGPL 1310
824 DSNKKT000SLPPPPPSLDOP--QPMEOGSS--GVLGKREB0GK--PERSSSP-- 873
1311 -SPPTKENASAMPVAVATM0EPATAPSTNPSLSLKTMI0AEVYKSLKEVASSPFKD 1369
874 -ASALTPEGE--ATSVT-----LVEELSLQE----- 896
1370 ESQDLVPAKLVNMTSQMNGFNSTFRPSDVANHMEQSDDEGLVASGSSESNNSGTEV 1429
897 --AMRKEPGESS-----SRKACEVGOAFPSQAALDEHOKTHKEGPLFTCVFCROGF 947
1430 TSSSSSEBPKKSAVSLAPNRVSCPYQ0RMPFWSSLSLRHILTHGQ0K-FKSH0CLLFF 1488
948 LERATLKRMHLAH0Y0 965
1489 TTKSCNDRHLRKHGNAE 1506
RESULT 39
077275 PRELIMINARY; PRT; 1891 AA.
077275;
01-NOV-1998 (Tremblrel. 08, Created)
01-NOV-1998 (Tremblrel. 08, last sequence update)
01-MAR-2002 (Tremblrel. 20, last annotation update)
EG:66A1.1 protein.
PEB OR EG:66A1.1 OR CG12212.
Drosophila melanogaster (Fruit fly).
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A.
RA Ferraz C., Vidal S., Brun C., Bucheton A., Demaille J.G.;
RT "Sequencing the distal X chromosome of Drosophila melanogaster."
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
CC -i- SUBCELLULAR LOCATION: NUCLEAR (By similarity).
DR EMBL; AL031227; CAA20227.1; -.
DR HSSP; P07248; 2ADR.
DR FlyBase; FBgn003053; Peb.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 12.
DR PRINTS; PR00048; ZNCFINGER.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 10.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 1891 AA; 205366 MW; 9E882364C36BB9BF CRC64;

Query Match 5.8%; Score 308; DB 5; Length 1891;
Best Local Similarity 20.3%; Pred. No. 1,2e-09;
Matches 247; Conservative 130; Mismatches 425; Indels 416; Gaps 56;

QY 53 ACSTDPVWVITIGG0ENPNSSASSEPREGHN-NPOVMDTEHSNPDSGSSVPTDPTWG 111
DB 397 SCSTD-----VSSGSHSSKSSSSSLNNNNNSHKRANNLIKDE-----ELEVTEDODTEN 446
QY 112 PERGEESGHFLV-----AATGTAGGGGGLILASPKIGATPLPPPS 154
DB 447 KORRLKTTINNIIIESEQEDMDDEADADAVAMLTSTPOVATILLAGASASGA-----A 500
QY 155 TPAPPPPPPPPP-----PGVSG-----HNTPILIEELRYL--QQR 190
DB 501 SRSPTPSPASAPALLISCACGASDFETPLALCVNLDAMHSDIPKCKDC0EVIATNR0L 560
QY 191 QI0HOMQEOICR-OVLILGSLQ0VGAAPAS--SELPGTASTSKPLPLPSPIKPVQ 247
DB 561 QSH-----CRLNMLALAGLPLRLGASSSPLINEPEDEH0GDEL-----E 603
QY 248 TSKTLASSSSSSSSGAETPKQAFPHLYHPLDSQHPFSAGVGRSHKPTP----- 298
DB 604 QKERLASQSED-----FFHQLYLKHKTANCGAISHPPSPKHNPRADTK 647
QY 299 -----APSPALPGSTDLITASPLAPFPTTGILAAQCGAARGLEATASPG- 345
DB 648 DLADIQSTILNMTSSSSSPLRNEQ0SVNTPNS0YSLDG-----RD0EEAD0AFT 697
QY 346 -----LKPKN0SGELSYGEVMGPLEKRGRRHKRCFCAKVFGS 382
DB 698 SEPRRMKLRGEPPCKLCTAVFPLRLALKGNRVHVGAV-----GPAGPRFCMMCPYAVCD 752
QY 383 DSAIQIHLRSHTGERPYKCNVCGNRFTRGNLKVHF-HRH-----REK----- 424
DB 753 KAALVHRMRTNHGDRPEYECAVCNVAFPTTKANCERHLNRNHGRTSREVKRAIYHRAEDA 812
QY 425 -----YPHVQNNPHRPV-----EHLDVYITSSGLPYGMSVPEKA 459
DB 813 GCEDSKSRKLGEDLADTSFRTSISPTPPRPVNESKSQLKHM--LLGSHNLAPVNOQRPPLKI 870
QY 460 EEAATPGGVERKPLV-----ASTTALS-ATESLTLTLLSTAGTATADGLPAF 506
DB 871 QVKSIDQ--LVDKKPSAPAPQ00000000EKSGLALDFSMVDLDSKRP0TGAS----- 921
QY 507 NKFLVLMKAVEPKNADEPTPPSGESA-ISGVASSATLMO-----LSKIMTSL 555
DB 922 ---LTPAVPTTPPAVAVPYTPGGVGTPLDLAALIEQ00LLA00QLFGAGGEYMOQLFRSL 977
QY 556 PSMALTNHFRS-TGSFP-LPLCARALGASPSETSKLOLVEKIDROGAVAVTSAASGAP 613

D	b	978	-----MFOQSTSGFPFFPFMA---PPPOANEKPRMVSPPRNINMPV-GVGVGVP	1025
Q	y	614	TTSAAPSSSSASGPNOCVLCRLVLSCPRALRLHVGQHGEPFKCKVCGRAPSTGNLR	673
D	b	1026	V-----PFGCPKMYKLNKLVL-MPKOKORRT---EPPFACEHCSARTLASMME	1072
Q	y	674	AHF-----VGH-----KASPARA-----	687
D	b	1073	RHVQKHPOFYAORORSGHHVMRGASNVAAAAAAATVMAGPBGSSGFGSNH	1132
Q	y	688	-----QNSCPICQ---KFTNAVTLOQHR-----MLHGQ	715
D	b	1133	HGHGHSCHSHGARIPSEYKVCALIAOQLKANKNTDLQALAHGSSVAGNPLHFGYP	1192
Q	y	716	IPNGGALPEGGGAOENGSEOSTVSGASF---POOQOQPPPEELSEEEDEDEE	771
D	b	1193	LTNPS---FMHNSSGQNGGATATLDDDEPKLIDEDENHDEVEDVDDEDEDEE	1249
Q	y	772	EEDVTDSDSL-----AGRGSESGE---KAISVRGDSSEBAGSEEVGTVAATAAGKEM	823
D	b	1250	MEDEPEPELILDEQPAKEBAEEQELPKPLEQLGTEKAAQKMAETI--LEQAIKAGKPL	1307
Q	y	824	DSNEKTQOSSLRPPRRDLDOP---QPMEOGSS--GYLGKKEEGK--PERSSSP---	873
D	b	1308	-SPPTKEKASPNAPVTVAATMQEPATITAPSTNPSLKTMTIAQAEYVKSLKEVASSPKD	1366
Q	y	874	-ASALTPEGE---ATSVT-----LVEELSLQE-----	896
D	b	1367	ESQDLVPVAKLVNATNSQNNGFNSYFRPSDVANHMEQSDSEELVAGSGSESNNSGTE	1426
Q	y	897	---AMRKEPGESS-----SRKACEVCGQAFRPSQALAEHOKTHPEKGLFTCVCFROGF	947
D	b	1427	TSSSSSEPKKKAAYSLAPKRVSCPYCQRMFPMWSSSLRHILTHTGOKP-FKCSHCLLF	1485
Q	y	948	LERATLKKHMLLHHQVQ	965
D	b	1486	TTKSNCDRHLLRKHNVE	1503
RESULT 40				
ID	090783	PRELIMINARY;	PRT:	1173 AA.
AC	090783;			
DT	01-NOV-1996	(Tremblrel. 01, Created)		
DT	01-NOV-1996	(Tremblrel. 01, Last sequence update)		
DT	01-MAR-2002	(Tremblrel. 20, Last annotation update)		
DE	Prockir2 (Fragment).			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BONE MARROW CELLS;			
RA	MEDLINE=95047430; PubMed=7958949;			
RT	Schuetz B.; Niessing J.;			
RT	"Cloning and structure of a chicken zinc finger cDNA:restricted			
RL	expression in developing neural crest cells.";			
RL	Gene 148:227-284(1994).			
DR	EMBL; X56805; CAA40140.1; .			
DR	HSSP; P08047; 1SP2.			
DR	InterPro: IPR000832; znf_C2H2.			
DR	Pfam: PF00966; znf_C2H2; 16.			
DR	SMART; SM00355; znf_C2H2; 16.			
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.			
DR	PROSITE; PS00157; ZINC_FINGER_C2H2_2; 15.			
KV	DNA-binding; Metal-binding; Zinc-finger.			
FT	NON_TER			
FT	SEQUENCE	1173 AA; 127327 MW; 493BF4C07300FA10 CRC64;		
Query Match				
5.8%; Score 307; DB 13; Length 1173;				

Best Local Similarity 21.4%; Pred. No. 7.6e-10;				
Matches 189; Conservative 90; Mismatches 314; Indels 292; Gaps 36;				
Q	y	292	RSKHPTPAPSPALPGSDQLASPHLARPTSTGLAAOQLGAARGLAENTSPGLLRKNG	351
D	b	9	RPYKSECPKAFKRSLSLIQHVSNA-----TGLRFKC--ALCGLAF-----KW	50
Q	y	352	SGELSYGEVMGPRLEKPGGRHNRKCFCAKVFYGSALQILHRSHTGERPKYKNCVGNREPTR	411
D	b	51	SSHXYHILRHQHTGERP---YKCTSCPFAFNSSSLRRHHIHTGERPYVSACGKATQS	107
Q	y	412	GNLKVHNRHREKRYHYQMNPRVPEHLDYITSSG--LPYGNVPRPEKAEERATPGG	469
D	b	108	TNLRHQHTHTGERPY-----ACSHCSKPTTHSSNLLHQRTHSSTRSHKCPACPKAF	160
Q	y	470	VERKPLVASTALSTATESLTLLSTAGTATAGLPAFNKFLMKATPERKKADENT----	525
D	b	161	VSDACLQKHILQSHNASPLLPSPSPQLSPPP-----LLEAVEMLYKTECTYLSP	212
Q	y	526	-----PPGSESAISGVAESSTATLMQLSKMTSLPSWMLTNHFKSTGS--	570
D	b	213	COBQLLGHQSHRQPRATPGD-VTATATHTRCPTCGKTFKNSSG-----LARHRHSHAER	265
Q	y	571	-FPLPLCARALGASPSSETSKLQQLVEKIDROGAVAVTSAAGAPTSAPAPSSASSGPN	629
D	b	266	RYKSCQCHRSFG-----QLAGLLGHORG---HSAETPHRPPTPATPTPTSVSERPY	312
Q	y	630	QCVIY-----	642
D	b	313	QCTECGKAFKSGSLGRYHMRDHTGERPYKSECPKAFKRSLLAHQVHTGLKAYKCP	372
Q	y	643	A-----LRLYGO-----HGERPFPCKYCGRA	665
D	b	373	CGLTFKSSNHYHLRLHRTGERPRYKCPDCKPAFNKTSCLGRPHQLHTGERPHACPIGKA	432
Q	y	666	FSTRGNLRAR--FVG-----HKASPARAQNSCPICOKKFTNA	701
D	b	433	FTQTSNLRHQHTHTGERPYACSHCGKTFTHSSNLQLRQTHSSARPHQCPKAFVMA	492
Q	y	702	VTLQAHYRMHLGGQIPNGTGALPREGGAQENG-----SEBSTYSGASGSPQOQSOQ-	753
D	b	493	SYLQRHLRTNAAG--PKGS---PRPALTPRODGPVYLQALSLLEVTAADAHFTLLQTPQG	547
Q	y	754	-----PSPPE--ELSEEEEEDEEEDVDYDSDSL--AGRGS-----	786
D	b	548	LQILPSPPAPQKLLLPATQPPPKHOGESPTPGSLLVPSTGTLPLRLQAVNAV	607
Q	y	787	ESGGEKAIYVRGDSSE--ASGAEEEVGTAAATAGKEMDSNEKT--TOQSSLRPPPPPD	842
D	b	608	QGTGTGLIVLQGLPEQRLHPAGIPIHGQAAVEGAAY---RLQANEVTVNVLQALPQPDVT	664
Q	y	843	SLD-----QPPMGOSSGVLGKKEGGKPERSSSPASALTPREGENTSVTLVEELS	893
D	b	665	SIPLOATDMTNVLQALPQ-----PTDVTNLIQALTEVTNVO---	701
Q	y	894	LOEAMRKEPGESSSRKACEVCGQAFRPSQALAEHOKTHPEKGLFTCVCFROGLELATL	953
D	b	702	LQALLQPTDVTNIOQATEVANVLQ--QALSQPDVTNIE-----LQAAEV	745
Q	y	954	KKHMLLHHQVQPARPHQPNITIALSLVPGCSPTSTGSLSPPR	998
D	b	746	TN-----AHLQVLP---QPSVEVTNIQLDQG---EVTNVQQLQTLPQ	779

Search completed: January 13, 2003, 15:22:30
Job time : 73.0643 secs

OM protein - protein search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

Run on: January 13, 2003, 15:17:40 ; Search time 37.9432 Seconds
(without alignments)
3518.868 Million cell updates/sec

Title: US-09-988-117-3

Perfect score: 5301
Sequence: 1 MAOETGSSSRLLGPGCEPAE.....SSIPSPGLSPFPKKDPTMP 1002

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : A.Geneseq.101002:*

1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5301	100.0	1002	23	AAE17955
2	4677.5	88.2	1005	23	AAE17954
3	4672.5	88.1	1005	23	AAE17967
4	4669.5	88.1	1005	23	AAE17968
5	4625.5	87.3	1019	22	AAW79822
6	4615.5	87.1	1007	22	AAW78838
7	845.5	15.9	813	22	AAW78838
8	765	14.4	1373	22	ABW62368
9	736.5	13.9	1267	22	ABW61050
10	697	13.1	172	22	AAO01962

11	529	10.0	200	22	AAW92796
12	482.5	9.1	330	22	ABB30894
13	482.5	9.1	330	22	ABB36076
14	482.5	9.1	330	22	AAW56858
15	482.5	9.1	330	22	AAW69245
16	482.5	9.1	330	22	AAW17074
17	482.5	9.1	330	22	AAW29568
18	482.5	9.1	330	22	AAW04775
19	482.5	9.1	330	23	ABG38855
20	482.5	8.3	336	22	ABW32272
21	482.5	8.3	336	22	ABW37530
22	482.5	8.3	336	22	AAW58186
23	438.5	8.3	336	22	AAW70641
24	438.5	8.3	336	22	AAW18485
25	438.5	8.3	336	22	AAW30956
26	438.5	8.3	336	22	AAW06075
27	438.5	8.3	336	23	ABG40328
28	337	6.4	927	22	AAW80283
29	330.5	6.2	619	22	AAW79299
30	324.5	6.1	869	22	AAW95103
31	322.5	6.1	702	23	ABW04333
32	321.5	6.1	498	22	AAW94659
33	320.5	6.0	799	21	AAW21003
34	311.5	5.9	744	22	ABW59449
35	311.5	5.9	744	22	ABW66002
36	311.5	5.9	744	22	ABW66572
37	309	5.8	934	22	ABW69958
38	309	5.8	1893	22	ABW59829
39	305.5	5.8	1196	22	ABW28743
40	304	5.7	725	22	AAW95116
41	303	5.7	722	22	AAW39029
42	300.5	5.7	823	22	AAW9340
43	299	5.6	894	22	AAW99335
44	297.5	5.6	515	22	ABW31309
45	297.5	5.6	515	22	ABW36515
46	297.5	5.6	515	22	AAW21854
47	297.5	5.6	515	22	AAW57278
48	297.5	5.6	515	22	AAW69683
49	297.5	5.6	515	22	AAW17495
50	297.5	5.6	515	22	AAW50016
51	297.5	5.6	515	22	AAW05170
52	297.5	5.6	515	23	ABW39299
53	297.5	5.6	890	22	AAW99339
54	292.5	5.5	812	22	AAW99337
55	292.5	5.5	624	23	ABW57324
56	290	5.5	883	23	ABW22743
57	290	5.5	1185	20	AAW3497
58	289.5	5.5	567	22	ABW16854
59	289.5	5.5	567	22	AAW38658
60	288.5	5.4	880	22	AAW99336
61	288.5	5.4	884	22	AAW99334
62	288.5	5.3	2639	22	AAW15016
63	282.5	5.3	567	21	AAW93988
64	281	5.3	882	22	ABW06465
65	281	5.3	1212	22	ABW00399
66	281	5.3	1472	22	AAW31578
67	279.5	5.3	491	21	AAW52153
68	278.5	5.3	1230	22	AAW30831
69	277.5	5.2	582	23	ABW7429
70	277.5	5.2	832	22	ABW10543
71	277.5	5.2	2703	22	ABW63299
72	276.5	5.2	675	22	AAW94388
73	273.5	5.2	556	22	ABW60575
74	272.5	5.1	701	22	AAW99341
75	270.5	5.1	577	22	AAW15930
76	270.5	5.1	839	21	AAW43281
77	269.5	5.1	756	22	ABW08429
78	269.5	5.1	839	22	AAW93280
79	269.5	5.1	1520	22	ABW27130
80	269	5.1	691	22	AAW99338
81	269	5.1	810	21	AAW73351
82	269	5.1	1054	22	ABW57993
83	268.5	5.1	714	22	ABW17368

Human protein sequ
Peptide #3545 enco
Peptide #3582 enco
Human brain expres
Human bone marrow
Peptide #3508 enco
Peptide #3605 enco
Peptide #3457 enco
Human peptide enco
Peptide #4923 enco
Peptide #5036 enco
Human brain expres
Human bone marrow
Peptide #4919 enco
Peptide #4993 enco
Peptide #4757 enco
Human peptide enco
Human protein SEQ
Human protein SEQ
Human protein SEQ
Human zinc finger
Human protein sequ
Human nucleic acid
Drosophila melanog
Drosophila melanog
Drosophila melanog
Drosophila melanog
Novel human diagno
Novel human diagno
Human protein sequ
Drosophila melanog
Human Rtl1 gamma p
Human novel secret
Human OREF ORF3045
Novel human diagno
Human protein sequ
Novel human diagno
Mouse Rtl1 gamma p
HTRM clone 1484257
Drosophila melanog
Novel human diagno

84	268.5	5.1	1252	22	AAW79739	Human protein SEO
85	268	5.1	1205	22	ABG05068	Novel human diagno
86	268	5.1	1205	22	ABG15312	Novel human diagno
87	268	5.1	1214	23	ABG27121	Novel human diagno
88	267.5	5.0	1267	22	ABB97532	Novel human protei
89	267.5	5.0	820	22	ABB95565	Drosophila melanog
90	267	5.0	613	22	ABB95862	Human protein sequ
91	266.5	5.0	1083	22	ABB12422	Human bone marrow
92	266	5.0	481	22	ABB92770	Human protein sequ
93	266	5.0	481	23	ABB97545	Novel human protei
94	266	5.0	791	22	ABB67268	Drosophila melanog
95	266	5.0	845	22	ABB64280	Drosophila melanog
96	265	5.0	755	22	AAW40916	Human polypeptide
97	265	5.0	809	22	AAW38689	Human polypeptide
98	264	5.0	804	22	AAW41058	Human polypeptide
99	263.5	5.0	529	22	AAW93633	Human polypeptide,
100	263	5.0	803	22	AAW92722	Human polypeptide

ALIGNMENTS

RESULT 1

AAE17955 standard; Protein; 1002 AA.

AAE17955;

07-MAY-2002 (first entry)

Mouse Sal2 protein.

Mouse; proliferative disorder; tumor host range mutant virus; cancer;

T-HR mutant; Sal2 protein; ovarian tumour.

Mus musculus.

Key Location/Qualifiers

Misc-difference 252 /note= "Encoded by CTG"

WO200204596-A2.

17-JAN-2002.

05-JUL-2001; 2001WO-US21354.

07-JUL-2000; 2000US-216723P.

19-MAR-2001; 2001US-0812471.

19-MAR-2001; 2001US-0812633.

(HARD) HARVARD COLLEGE.

(BGM) BRIGHAM & WOMENS HOSPITAL INC.

Benjamin TL, Li D, Mok SC, Cramer DW, Ma Y;

WPI: 2002-164637/21.

N-PSDB; AAD28652.

Detecting protein involved in susceptibility to proliferative disease, by infecting normal and abnormal proliferating cells with mutant virus, detecting mutated protein allowing growth of mutant on abnormal cells

Disclosure; Page 85-87; 92pp; English.

The invention relates to a method for the identification of genes and their encoded proteins involved in susceptibility to proliferative disorders, including cancer using a tumor host range mutant (T-HR mutant) virus. The invention also provides the use of Sal2 genes and proteins in methods of identifying a mammal having, or at a risk of acquiring a proliferative disease. T-HR mutants are used to kill cancer cells such as one carrying a Sal2 alteration. Transgenic and knockout

CC mouse comprising Sal2 nucleic acid are useful as research tools to determine genetic and physiological features of cancer and for CC identifying compounds that can affect ovarian and other tumours. The CC present sequence is mouse Sal2 protein.

Sequence 1002 AA:

Query Match 100.0%; Score 5301; DB 23; Length 1002; Best Local Similarity 100.0%; Pred. No. 0; Matches 1002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAQETGSSRLGGCGCPAERGDAEEHHPOVCAKCAQFSPPTFLAQNQSCCTPPV	60
DB	1	MAQETGSSRLGGCGCPAERGDAEEHHPOVCAKCAQFSDPTFLAQNQSCCTPPV	60
QY	61	MTIIGQENPSNSASSAPRPGHRSQVMDTEHNPDPGSSGPPDPMTGPRRGESS	120
DB	61	MTIIGQENPSNSASSAPRPGHRSQVMDTEHNPDPGSSGPPDPMTGPRRGESS	120
QY	121	GQFLVATGTAAAGGGGLILASPKLGATPLPESTPAPPPPPPPPGVSGHLNPLT	180
DB	121	GQFLVATGTAAAGGGGLILASPKLGATPLPESTPAPPPPPPPPGVSGHLNPLT	180
QY	181	LEELRVLQQRQIHQMOMTEQICROVLLIGTGQTVGAPSPSELPTGAASSTKPLPLF	240
DB	181	LEELRVLQQRQIHQMOMTEQICROVLLIGTGQTVGAPSPSELPTGAASSTKPLPLF	240
QY	241	SPKPAQQTGTTASSSSSSSSGAEPKQAFPHLYHPLGSGHFVSGVGRSHKPPAPS	300
DB	241	SPKPAQQTGTTASSSSSSSSGAEPKQAFPHLYHPLGSGHFVSGVGRSHKPPAPS	300
QY	301	PALPGSTDOLIASPHLAFPTGTGLLAOCGAARGLEAAASPGLLKPKNGSGELGGEVY	360
DB	301	PALPGSTDOLIASPHLAFPTGTGLLAOCGAARGLEAAASPGLLKPKNGSGELGGEVY	360
QY	361	SSLEKPGRRHKRCFAKVFSDSALQIHLRSHTGERRPKYCNVCGNRTFTRGNLKVPHRH	420
DB	361	SSLEKPGRRHKRCFAKVFSDSALQIHLRSHTGERRPKYCNVCGNRTFTRGNLKVPHRH	420
QY	421	REKYPHVQNNPVPVPHLDVYITSSGLPYGMSVPEPKAEEAGTGGVERKPLVASTTA	480
DB	421	REKYPHVQNNPVPVPHLDVYITSSGLPYGMSVPEPKAEEAGTGGVERKPLVASTTA	480
QY	481	ISATESLTLTSTGTAFAVAGLPTFNKFLMKAVEPKSKABENPPGSDGSAIAGADSG	540
DB	481	ISATESLTLTSTGTAFAVAGLPTFNKFLMKAVEPKSKABENPPGSDGSAIAGADSG	540
QY	541	SATRMQLSKLVTSIPSWALLTNHLKSTGSEFPFVYLEPLGASPSFTSKLQQLVEKIDROG	600
DB	541	SATRMQLSKLVTSIPSWALLTNHLKSTGSEFPFVYLEPLGASPSFTSKLQQLVEKIDROG	600
QY	601	AVAVASTAGAPTTSPAPSSASGPNOCYICLRVLSCPRALRLHYGHGGRPRPKCYC	660
DB	601	AVAVASTAGAPTTSPAPSSASGPNOCYICLRVLSCPRALRLHYGHGGRPRPKCYC	660
QY	661	GRAFTRGTLRAHFVGHKTSPPARAQNSCPICQKFTNAVYLQOHVHRNHLGGQINGGSA	720
DB	661	GRAFTRGTLRAHFVGHKTSPPARAQNSCPICQKFTNAVYLQOHVHRNHLGGQINGGSA	720
QY	721	ISEGGAQOENSSEOSTAGSPGSPPOQOQPSPEEMSEEEDEEEDVTDBSLAG	780
DB	721	ISEGGAQOENSSEOSTAGSPGSPPOQOQPSPEEMSEEEDEEEDVTDBSLAG	780
QY	781	RGSSEGGKATSVRGDSPEVSGAEEEVATVAPPTYKEDNSNKAQOHTLPPPPPPDN	840
DB	781	RGSSEGGKATSVRGDSPEVSGAEEEVATVAPPTYKEDNSNKAQOHTLPPPPPPDN	840
QY	841	LDHPQPMQGTSDVSGAMEEAKLEGISPMALTOEGESTPLVEELNPEAMKRPD	900
DB	841	LDHPQPMQGTSDVSGAMEEAKLEGISPMALTOEGESTPLVEELNPEAMKRPD	900
QY	901	ESSRRKACEVCGSPPTOTALAEHQTHRKDGLFVCFCRGGFLDRATLKKHMLAHQ	960
DB	901	ESSRRKACEVCGSPPTOTALAEHQTHRKDGLFVCFCRGGFLDRATLKKHMLAHQ	960

Db	901	ESSSRKACEVCGSPPTQALAEHQHTHPKDGFLFTCVCRQGFGLDRATLKKHMLAHNQ	960
QY	961	VPPPAHPGPNATLTLVPGCSSIPSPGSLPPPRKDDPTMP	1002
Db	961	VPPPAHPGPNATLTLVPGCSSIPSPGSLPPPRKDDPTMP	1002
RESULT 2			
AAEL17954	standard: Protein; 1005 AA.		
AAEL17954			
AAEL17954			
07-MAY-2002	(first entry)		
Human Sal2 protein.			
Human	proliferative disorder; tumor host range mutant virus; cancer;		
T-HR mutant; Sal2 protein; ovarian tumor; chromosome 14q12.			
OS	Homo sapiens.		
Key	Location/Qualifiers		
Misc-difference	15	/note= "Encoded by TGC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"	
Misc-difference	22..23	/note= "Encoded by bases 4611-5022 of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"	
Misc-difference	120	/note= "Encoded by CCA of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"	
Misc-difference	545	/note= "Encoded by CGC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"	
Misc-difference	552	/note= "Encoded by GTG of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"	
Misc-difference	573	/note= "Encoded by TTC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"	
Misc-difference	575	/note= "Encoded by TAT of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"	
Misc-difference	576	/note= "Encoded by GTG of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"	
Misc-difference	577	/note= "Encoded by ATC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"	
Misc-difference	578	/note= "Encoded by GAG of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"	
Misc-difference	579	/note= "Encoded by CCC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"	
WO200204596-A2.			
17-JAN-2002.			
05-JUL-2001.	2001MO-U\$21354.		
07-JUL-2000.	2000US-216723P.		
19-MAR-2001.	2001US-0812471.		
19-MAR-2001.	2001US-0812633.		
(HARD) HARVARD COLLEGE.			
(BGHM) BRIGHAM & WOMENS HOSPITAL INC.			
Benjamin TL, LI D, Mok SC, Cramer DW, Ma Y;			

Query Match	Best Local Similarity	88.28;	Score 4677.5;	DB 23;	Length 1005;
Matches 894;	Conservative 25;	Mismatches 83;	Indels 3;	Gaps 3	
1 MAOEGRSSSLRGPCGPAERGGASSEHHNPVCAKCAQFSPTEFLAHONSCCTDPV	1 MHESERSRLGYPAGEPAELGDASEEDHPVCAKCAQFTPTPEFLAHONACSTDPPV	1	1	1	1
61 MYIGGOENPNSSASSAPREGHRSRQVMDTEHNSNPDSGSSGPDPTWGPERRGESS	61 MYIGGOENPNSSASSAPREGHRSRQVMDTEHNSNPDSGSSGPDPTWGPERRGESS	61	61	61	61
121 GQFLVAATGTAAGGGGGLTASPKLGARPLPESPPAPPPPPPPPPPPPPPPPPPPPPPP	121 GQFLVAATGTAAGGGGGLTASPKLGARPLPESPPAPPPPPPPPPPPPPPPPPPPPPPP	121	121	121	121
121 GHFLVATGTAAGGGGGLTASPKLGARPLPESPPAPPPPPPPPPPPPPPPPPPPPPPP	121 GHFLVATGTAAGGGGGLTASPKLGARPLPESPPAPPPPPPPPPPPPPPPPPPPPPPP	121	121	121	121
181 LELRLVLOQRIHOMOMTEQICRQVLLGSLQTVGAPASPSSELPGTAASSTKPLPLE	181 LELRLVLOQRIHOMOMTEQICRQVLLGSLQTVGAPASPSSELPGTAASSTKPLPLE	181	181	181	181
181 LELRLVLOQRIHOMOMTEQICRQVLLGSLQTVGAPASPSSELPGTAASSTKPLPLE	181 LELRLVLOQRIHOMOMTEQICRQVLLGSLQTVGAPASPSSELPGTAASSTKPLPLE	181	181	181	181
241 SPIKPAOTGTTA-SSSSSSSSSGAEPKQAFHLNHPGSGQHFVSQGVGRSHKPTAP	241 SPIKPAOTGTTA-SSSSSSSSSGAEPKQAFHLNHPGSGQHFVSQGVGRSHKPTAP	241	241	241	241
241 SPIKPAOTGTTA-SSSSSSSSSGAEPKQAFHLNHPGSGQHFVSQGVGRSHKPTAP	241 SPIKPAOTGTTA-SSSSSSSSSGAEPKQAFHLNHPGSGQHFVSQGVGRSHKPTAP	241	241	241	241
300 SPALPGSTDOIASPPLHAFPGTTGLLAACLGARGLAEMAASPGLLPKNGSGELGIGYEV	300 SPALPGSTDOIASPPLHAFPGTTGLLAACLGARGLAEMAASPGLLPKNGSGELGIGYEV	300	300	300	300
301 SPALPGSTDOIASPPLHAFPGTTGLLAACLGARGLAEMAASPGLLPKNGSGELGIGYEV	301 SPALPGSTDOIASPPLHAFPGTTGLLAACLGARGLAEMAASPGLLPKNGSGELGIGYEV	301	301	301	301
360 ISSLEKGGRRHKRCFCACVFGSDAQLIHLRSHTGERPYKCNVCGNRRFTTNGMLKVHFR	360 ISSLEKGGRRHKRCFCACVFGSDAQLIHLRSHTGERPYKCNVCGNRRFTTNGMLKVHFR	360	360	360	360
361 MGPLEKPGRRHKRCFCACVFGSDAQLIHLRSHTGERPYKCNVCGNRRFTTNGMLKVHFR	361 MGPLEKPGRRHKRCFCACVFGSDAQLIHLRSHTGERPYKCNVCGNRRFTTNGMLKVHFR	361	361	361	361
420 HREKYPVQNMHPHYPEHLDVYITSSGLRYGMSVPPKAEDEAATPGGVERKPLVASTT	420 HREKYPVQNMHPHYPEHLDVYITSSGLRYGMSVPPKAEDEAATPGGVERKPLVASTT	420	420	420	420
421 HREKYPVQNMHPHYPEHLDVYITSSGLRYGMSVPPKAEDEAATPGGVERKPLVASTT	421 HREKYPVQNMHPHYPEHLDVYITSSGLRYGMSVPPKAEDEAATPGGVERKPLVASTT	421	421	421	421
480 ALSTESLTLSTGTAIVAPLPTFNKFLVLMKAVEKSKADENTPPGSEGAIAGVADS	480 ALSTESLTLSTGTAIVAPLPTFNKFLVLMKAVEKSKADENTPPGSEGAIAGVADS	480	480	480	480
481 ALSTESLTLSTGTAIVAPLPTFNKFLVLMKAVEKSKADENTPPGSEGAIAGVADS	481 ALSTESLTLSTGTAIVAPLPTFNKFLVLMKAVEKSKADENTPPGSEGAIAGVADS	481	481	481	481
540 GSATRMQSLKVLTSPLSMALTNHLKSTGSEPPRYVLEPLGASPSSESKQDLVETIDQ	540 GSATRMQSLKVLTSPLSMALTNHLKSTGSEPPRYVLEPLGASPSSESKQDLVETIDQ	540	540	540	540
541 STAILMOLSKLMTSLPSSWALTNHFKSTGSEPPRLCARALGASBSETSKQDLVETIDQ	541 STAILMOLSKLMTSLPSSWALTNHFKSTGSEPPRLCARALGASBSETSKQDLVETIDQ	541	541	541	541
600 GAVAVASTASCAPTTSAPRPSA-SSGPNQVCICRLVLSCPRALRLAHGQHGGERPFCK	600 GAVAVASTASCAPTTSAPRPSA-SSGPNQVCICRLVLSCPRALRLAHGQHGGERPFCK	600	600	600	600
601 GAVAVASTASCAPTTSAPRPSA-SSGPNQVCICRLVLSCPRALRLAHGQHGGERPFCK	601 GAVAVASTASCAPTTSAPRPSA-SSGPNQVCICRLVLSCPRALRLAHGQHGGERPFCK	601	601	601	601

IY	659	VCGRAFSRGNLRAHFVHGKTKSPAAANAQNSCPICCKKFTTAVTLQOHVRHMLGQIDPNCG	718
JD	661	VCGRAFSTRGNLRAHFVGHKASPPAARAQNSCPICCKKFTTAVTLQOHVRHMLGQIDPNCG	720
IY	719	SALSSGGGAQAQENSSSEQSTVASGPSPPOSQQPPEEEMS-EEEEDEEEEDVTDEDSDS	777
JD	721	TALPFGGGAQAQENGSGSVQSAGSPFPQQSQQPSPEELSEEDEEDEEBEVTEDEDS	780
IY	778	LARGSESGGEKAISVRGDSEEVSCAAEEVANISVAAPTVKEKDNSNEKADQHTLPPLPPP	837
JD	781	LARGSESGGEKAISVRGDSEEAASAEEEGVTAATAATACKENDSNKKTQQSSLPPLPPP	840
IY	838	PDNLDHPQMEGSTDVSACAMEEAEAKLEGISSPMALLTOEGESTPTLYBELNLPKAMK	897
JD	841	PDSDLDPQPMRGSSGVGLGKREGEKGKPESSSPASALTPEGEATSTVLVELSLQEMRK	900
IY	898	DPEGSSGRKACEVCGOSPPTQTALBEHQTHPKDGLFTFCVFCROGFLDRATLKHHLLA	957
JD	901	EPGESSSKKACEVCQAEPFOALAEHQTHPEKEGFLTFCVFCROGFLERATILKKHMLLA	960
IY	958	HNOVPFAPHGPNONTATLSLVGCSSIPSGISPPRRDDPTMP	1002
JD	961	HNQVQFAFHGPNONTALSILVGCSPSTITSGLSPPRRDKDPTIP	1005
<hr/>			
:RESULT 3			
AEI7967			
D	AAEI7967	standard; Protein; 1005 AA.	
X			
X	AAEI7967;		
X			
X	07-MAY-2002	(first entry)	
X			
E	Human Sal2 protein mutant (S73C).		
X			
X	Human; proliferative disorder: tumor host range mutant virus; cancer;		
X	T-HR mutant; Sal2 protein; ovarian tumour; mutant; mutlein.		
S	Homo sapiens.		
S	Synthetic.		
X			
H	Key	Location/Qualifiers	
T	Misc-difference 73	/note= "Wild-type Ser substituted with Cys"	
X			
N	WO200204596-A2.		
X			
X	17-JAN-2002.		
D			
X	05-JUL-2001; 2001WO-US21354.		
F			
X	07-JUL-2000; 2000US-216723P.		
R	19-MAR-2001; 2001US-0812471.		
R	19-MAR-2001; 2001US-0812633.		
A			
A	(HARD) HARVARD COLLEGE.		
A	(BGHM) BRIGHAM & WOMENS HOSPITAL INC.		
X			
I	Benjamin TL, Li D, Mok SC, Cramer DW, Ma Y:		
X			
R	WPI; 2002-164637/21.		
X			
X			
T	Detecting protein involved in susceptibility to proliferative disease,		
T	by infecting normal and abnormal proliferating cells with mutant virus		
T	detecting mutated protein allowing growth of mutant on abnormal cells		
T	-		
X			
S	Claim 35; Page -; 92pp; English.		
X			
X	The invention relates to a method for the identification of genes		
X	and their encoded proteins involved in susceptibility to proliferative		
X	disorders, including cancer using a tumor host range mutant (T-HR		
X	mutant) virus. The invention also provides the use of Sal2 genes and		

CC proteins in methods of identifying a mammal having, or at a risk of
CC acquiring a proliferative disease. T-HR mutants are used to kill cancer
CC cells such as one carrying a Sal2 alteration. Transgenic and knockout
CC mouse comprising Sal2 nucleic acid are useful as research tools to
CC determine genetic and physiological features of cancer and for
CC identifying compounds that can affect ovarian and other tumours. The
CC present sequence is human Sal2 protein mutant (573C).
CC Note: This sequence is not shown in the specification but is derived
CC from the human Sal2 wild-type protein shown as SEQ ID NO:1 (AAE17954)
CC in page 77-79 of the specification.

AA	Sequence	1005	AA:
SQ			

Query Match	88.1%;	Score 4672.5;	DB 23;	Length 1005;
-------------	--------	---------------	--------	--------------

Best Local Similarity 88.9%; Pred. NO. 4.0e-284;
Matches 893; Conservative 25; Mismatches 84; Indels 3; Gaps 3;

QY 1 MAQETGSSRLGPGCEPAERGGDASEEHHPQVCAKCAQFSDPTEFLAHQNSCCTDPV 60

DB 1 MAHESKSSRLGVPAGEPAELGGDASEEDHPQVCARCCAQFTDPTFEFLAHQNACTDPV 60

QY 61 MVIIGQENPSNSSASSAPRPEGHRSQVMDTEHSNPPDSGSSGPPDPTWGPERRGEES 120

Db 61 MIIIGQENPNNCSSASSEPRPEGHNPPQVMDTEHSNPPDSSSVPTDPTWGPERRGEES 120

QY 121 GQFLVAATGTAAGGGGCLILASPKLGATPLPEESTPAPPPPPPPPCVGSCHLNIPLI 180

Db 121 GHFLVAATGTAAGGGGLIASPKLGATPLPPESTPAPPPPPPPPPGVGSGHLNIPLI 180

QY 181 LEEIRVLQQRQIHQMOMTEQICRQVLLGSLGQTVGAPASPSELPGTGAASSTKPLPLF 240

Db 181 LEEELRVLOOQROIHMOMTEOICROVLLGSLGOTVGAPASPSELPGTASSTKPLPLF 240

OV 241 SPIKPAOTGKTTA-SSSSSSSSSSGAEPPKOAFFHTYHPICSHHPFSVGGVGRSHKPTDAP 299

```

db      241 SPTRPVOTSKTIASSSSSSSSGAEFBRKAFENIYHPICSQUPREACCVBPEUTEND 300
||||| |||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

[illegible]

— — — — —

[illegible][illegible]

DB 361 MGPLEKPGGRHKCRFCAKVFGSDALQIHLRSHTGERPYKCNVCGNRFYTRGNLKVHEHR 420

420 HREKYPHVQMNPHVPPEHLDVITSSGLPYGMSVPPEKAEFEAGTPGGVERKPLVASTT 479

Db 421 HREKYPHVQMNPHRVPREHLDYVITSSGLPYGMSVPPEKAEAEAAATPGGVERKKPLVASTT 480

QY 480 ALSATESLTLSTGTSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSEGSIAIGVADS 539

Db 481 ALSATESLTLSTAGTATAPGLPAFNKEVLMKAVEPKKADENTPPGSEGS AISGVAES 540

QY 540 GSATRMQLSLVTSIPSWALLTNHLKSTGSPFPYPVLEPLGASPSSETSKLQQLVEKIDRQ 599

Db 541 STATIMQLSKLMTSLPSWALLTNHFKSTGSPPLPCARALGASPSSETSKLQQLVEKIDRQ 600

QY 600 GAVAVASTASGAPITTSAPAPSSSA-SGPNOCVICLRVLSCPRALRLHYGONGGERPFCK 658

Db 601 GAVAYTSAASGAPTTAPAPSSASGPNOCYTCI RYI SCPRAI RI HYGOHGGERPEKCK 650

0v 659 YCGRAESTBCNI.PAHEVGHKSPABAPONSCBICOKKETNAVETCOHYBMUICCOTDVC 710

667 WCCRAECMBDPII PAIIEVICIIV3 CDA AD AOVOCSTCOVETIAIIEI COVIEIVUC COCTEIVOC 700

[illegible]

```

01          CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
      :||| ||||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

DD /Z1 IALPEGGAQENGSEQSTVSGAGSEFQQQSQQPSPEEELSEEEEEDEEEEEEDVILDEDS /80

0Y 7/8 LAGRSES6GEKAISVRGDSEEVSGAEFEVATSVAAPTTVKEMDSNEKAPQHTLPPPPP 837

Db 781 LAGRGSESGEKAISVRDSEEAASGAEEEVGTVAATAATAGKEMDSNEKTTQQSSLP PPP 840

[illegible]

Query Match	88.1%;	Score 4669.5;	DB 23;	Length 1005;
Best Local Similarity	88.9%;	Pred No. 7.1e-284;		
Matches	893;	Conservative	25;	Mismatches 84; Indels 3; Gaps 3;
QY	1	MAOETGSSRLRGPCGPEAERGGDASEEHHPOVCACCAOAFSDPTEFLAONSCSTDPPV	60	
QY	1	MAHESRSRLGPAEPALGDASEEDHPQVCAKCAOFTDTEFLAONASTDPV	60	
QY	61	MYITGOENPNSNSASAPRPEGHRSQVMDTEHSNPPDSSGSPRPDPWPERGEESS	120	
QY	61	MYITGOENPNSNSASASEPRPEGHNPQVMDTEHSNPPDSSSVPTDPWPERGEESS	120	
QY	121	GQFLVATGTAAGGGGGLLASPRIGATPLPEESTPARPPPPPPPPVSGHNLPI	180	
QY	121	GHLVATGTAAGGGGGLLASPRIGATPLPEESTPARPPPPPPPPVSGHNLPI	180	
QY	181	LEELRLVLOORQHOMOMTEQICRQVLLLSIGQVVGAPASSELPCTGAASSTRPLPLF	240	
QY	181	LEELRLVLOOQRIHOMOMTEQICRQVLLLSIGQVVGAPASSELPCTGAASSTRPLPLF	240	
QY	241	SPFKPQOTGKTTA-SSSSSSSSGAEPKPKOAFHLYHPLGSGQHPFSVGVGRSHKPTPA	299	
QY	241	SPFKPQOTSTTLASSSSSSSSGAEPKPKOAFHLYHPLGSGQHPFSVGVGRSHKPTPA	300	
QY	300	SPALPGSTDOLIASPHLAPETGTLAOCIGAARGLAASPGILKPKNSGELCYGV	359	
QY	301	SPALPGSTDOLIASPHLAFSTTGTLAOCIGAARGLAATSPGLKPKNSGELSYGV	360	
QY	360	ISSELRKGGHHKCFCKKVGSDSALOHLRSHNGERYKCVNCGNRPFTRNKLKVFHR	419	
QY	361	MGPLEKGGHKKCFCKKVGSDSALOHLRSHNGERYKCVNCGNRPFTRNKLKVFHR	420	
QY	420	HREKYPHVQNNPVRPHLDVYITSSGLPYGMSVPEKAEAEATGGGVERRPLVAST	479	
QY	421	HREKYPHVQNNPVRPHLDVYITSSGLPYGMSVPEKAEAEATPGGVEKRRPLVAST	480	
QY	480	ALSTESTLTLSTSTRAVARGLPENKFLVLMKAVEPSKADENTPPGSEGSADAVAS	539	
QY	481	ALSTESTLTLSTSAGATATAGLPENKFLVLMKAVEPKADENTPPGSEGSASIVAES	540	
QY	540	GSATRMOLSLVTSLSFWALLTNHLKTSGFPPVPLRPLGASPSSTKLQOLVEKIDRQ	599	
QY	541	STALIMQLSKMTLPSMALLTNHFKSTGSPRLPCARALGASPSSTKLQOLVEKIDRQ	600	
QY	600	GAVAVASTAGAPTTAPAPSSA-SGRNOCYICRLVSCPRALRLHYOHGGERFPCK	658	
QY	601	GAVAVTSASAGAPTTAPAPSSASGPNOCYICRLVSCPRALRLHYOHGGERFPCK	660	
QY	659	VCGRAFSRGLRAHFVHKTSPARAONSCITCKKFTNATYTLQOHVNMHLGQIPNG	718	
QY	661	VCGRAFSRGLRAHFVHKTSPARAONSCITCKKFTNATYTLQOHVNMHLGQIPNG	720	
QY	719	SALSEGGGAOENSSSESTAGSGPSFPOPOSOAPSDEEMS-EEEDEDEEEDVTDSDS	777	
QY	721	TALPEGGGAOENSGSESTAGSARFPOOOSQOAPSEEBELSEEEDEDEEEDVTDSDS	780	
QY	778	LAGGSEGGKALSYRDSSEFVGAEEFVAATVTKEMDSNEKKAPOHLLPPPP	837	
QY	781	LAGGSEGGKALSYRDSSEFVGAEEFVVAATAATGKEMDSNEKKTQOSSLPPPP	840	
QY	838	PDNLDPHPMEOGTSDVGVGAMEEBKLKEISSPMALTOEGECTSTPLVELNLPEAMK	897	
QY	841	PDNLDPHPMEOGTSDVGVGAMEEBKLKEISSPMALTOEGECTSTPLVELNLPEAMK	900	
QY	898	DPGESSGKACBVGQSEPTQALBEOHKTNRKDGFLPTCYCFRQGLDRAATLKKHMLA	957	
QY	901	EPGESSSRKACBVGQSEPTQALBEOHKTNRHEGFLPTCYCFRQGLDRAATLKKHMLA	960	
QY	958	HHOVPFAPHGQNTATLPLVGCSSISPSGLSPPRRKDDPPM 1002		
QY	961	HHOVPFAPHGQNTATLPLVGCSSISPSGLSPPRRKDDPPM 1005		

Query Match	Best local Similarity	883; Conservative	87.3%; Score 4625.5; DB 22; Length 1019; Mismatches 883; Mismatches 81; Indels 3; Gaps 3;
16	GGEAERGGDSEEHHPQ	CAKCCAKOFSDPFEFLAHONSCCTDPDPMVTITGGGENSENSSA	75
30	GPEASENGDAASEEDHPO	YCAKCCAKOFLDPFTFLAHONACSTDPDPMVTITGGGENSENSSA	89
76	SSAPRDEGHSRSDVMDTEHSPNPDSSGSSGPPDPWTGDERRGEESGQFLVATGTAAAGG		135

Db	90	SSNP	PREGHNNPQYMDI	EHNSRPDPSGSSVPLDPTWGPERRGESSCHFLVLAATGTAAGG	149
QY	136	GGIL	ASP	KIGATPLPEESTPAPRPPPPPPPPPPGVSGHLNIPILIEELRVLIQRIHOW	195
Db	150	GGIL	ASP	KIGATPLPEESTPAPRPPPPPPPPPPGVSGHLNIPILIEELRVLIQRIHOW	209
QY	196	QME	EOICRQVLLGSLGQIVGAP	SPSLPCTGMASSNRKPLILPSPIKRPAQTGKTTA-S	254
Db	210	QME	EOICRQVLLGSLGQIVGAP	SPSLPCTGMASSNRKPLILPSPIKRPAQTGKTTA-S	269
QY	255	SSSS	SSSSSGAEP	PPKQAFPHLYHPLGSOHPFESVGVGRSHKPTPAESPALPGSTDOLIASP	314
Db	270	SSSS	SSSSSGAEP	PPKQAFPHLYHPLGSOHPFESVGVGRSHKPTPAESPALPGSTDOLIASP	329
QY	315	HLAF	PQTGLLAQCCIGAAGLEA	ASBGLKPKNGSGELGVEYISLIERKPGRRHKRF	374
Db	330	HLAF	PQTGLLAQCCIGAAGLEA	ASBGLKPKNGSGELGVEYISLIERKPGRRHKRF	389
QY	375	CAK	VFSDSALQHLIRSH	TERPCKCNWCGRRFTTQGLKVFHFRHREKYPVHOMNPHV	434
Db	390	CAK	VFSDSALQHLIRSH	TERPCKCNWCGRRFTTQGLKVFHFRHREKYPVHOMNPHV	449
QY	435	PEHL	DVITSSGLPYGNSV	PEKAEBAATPGGVEERKPLVASTALATESLTLTLSTGT	494
Db	450	PEHL	DVITSSGLPYGNSV	PEKAEBAATPGGVEERKPLVASTALATESLTLTLSTGT	509
QY	495	STAV	AVAGLPLPENK	FVLMKAAVEPKSKADENTPGSEGSIAIADVANGSATRMLSTLYSL	554
Db	510	STAV	AVAGLPLPENK	FVLMKAAVEPKSKADENTPGSEGSIAIADVANGSATRMLSTLYSL	569
QY	555	PSMA	LLTNHLKSTGSE	PEFVYLPPLGASPSSTSKLOQIVLEKIDRGAAVAVASTAGAPTT	614
Db	570	PSMA	LLTNHLKSTGSE	PEFVYLPPLGASPSSTSKLOQIVLEKIDRGAAVAVASTAGAPTT	629
QY	615	SAP	APSSSA-SEPN	QCVICLRVYISCPRALRHYQHOGGERPKCKVCGRAPSTRGNLRAH	673
Db	630	SAP	APSSSA-SEPN	QCVICLRVYISCPRALRHYQHOGGERPKCKVCGRAPSTRGNLRAH	689
QY	674	FVGH	KTSPPARA	RQNSCPICQKFTNNATVLIQOHVHMLGGQIPNGSGALSEGGAQOENSS	733
Db	690	FVGH	KTSPPARA	RQNSCPICQKFTNNATVLIQOHVHMLGGQIPNGSGALSEGGAQOENSS	749
QY	734	EOS	TASGPGSFP	QOSQSPSEELSEEEDEEDVDEDSILGRGSESGEKAIS	792
Db	750	EOS	TASGPGSFP	QOSQSPSEELSEEEDEEDVDEDSILGRGSESGEKAIS	809
QY	793	VRG	DEEVS	GAEEVATVAAAPRTVEMDSNEKARQTLPPRPPRPNLDLHPQMEQTS	852
Db	810	VRG	DEEVS	GAEEVATVAAAPRTVEMDSNEKARQTLPPRPPRPNLDLHPQMEQTS	869
QY	853	DVS	GAEMEEBAK	LEGISPPNAALTOEPEGSTPLVEELINLPEAMKDPGESSGRKACEVCG	912
Db	870	DVS	GAEMEEBAK	LEGISPPNAALTOEPEGSTPLVEELINLPEAMKDPGESSGRKACEVCG	929
QY	913	QSF	PQTALAEHQ	KTHPKRGPLFTCYFCRQGFIDRLATLKKHMLLAHQVPPFAHPQNTI	972
Db	930	QSF	PQTALAEHQ	KTHPKRGPLFTCYFCRQGFIDRLATLKKHMLLAHQVPPFAHPQNTI	989
QY	973	ATLS	LVPGCSSS	IPSPGLSPRPKDDPTMP	1002
Db	990	ATLS	LVPGCSSS	IPSPGLSPRPKDDPTMP	1019
RESULT 6					
ID AAM78838 standard; Protein; 1007 AA.					
XX AAM78838;					
XX 06-NOV-2001 (first entry)					
XX Human protein SEQ ID NO 1500.					
XX					

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
PN MO200157190-A2.
PD 09-AUG-2001.
PF 05-FEB-2001; 2001WO-US04098.
PR 03-FEB-2000; 2000US-046914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AU, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
XX N-PSDB; AAK51971.
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 3790-3792; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAK78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAK80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1007 AA:
Query Match 87.1%; Score 4615.5; DB 22; Length 1007;
Best Local Similarity 89.0%; Pred. No. 1.7e-280;
Matches 881; Conservative 24; Mismatches 82; Indels 3; Gaps 3;

QY 255 SSSSSSSGAEPPKQAFHLHYHDLGSQLHPFSVGVGRSHKPTPAAPSALPGSTDOLIASP 314
DB 258 SSSSSSSGAELPCKQAFHLHYHDLGSQLHPFSVGVGRSHKPTPAAPSALPGSTDOLIASP 317
QY 315 HLAFPGTGLLAAQCGAARGLEATAAPSGLLKPKNGSGELGYEYISSLEKGRGKRCF 374
DB 318 HLAFPGTGLLAAQCGAARGLEATAAPSGLLKPKNGSGELGYEYISLEKGRGKRCF 377
QY 375 CAKVFGSDALQIHLSHNGERYKCNVCGNRFTTRGNLKVHFRHREKYPHYOMNPHY 434
DB 378 CAKVFGSDALQIHLSHNGERYKCNVCGNRFTTRGNLKVHFRHREKYPHYOMNPHY 437
QY 435 PEHLDYVITSSGLPYGMSVPEKAEEDAGTPGGVERKPLVASTALATESLTLSTGT 494
DB 438 PEHLDYVITSSGLPYGMSVPEKAEEDAGTPGGVERKPLVASTALATESLTLSTGT 497
QY 495 STAVAPGLPTFNKFLVLMKAVEPKSKADENTPPGSEGSALAGVADGSATRMQLSKVTSL 554
DB 498 GTATAPGLPAFNKFLVLMKAVEPKSKADENTPPGSEGSALISGVAESSTATLMQLSKVTSL 557
QY 555 PSMALLTNHLKSTGSPFPVYLEPLGASPSFTSKLOQLVEKIDROGAVAVASTAGAPT 614
DB 558 PSMALLTNHLKSTGSPFPVYLEPLGASPSFTSKLOQLVEKIDROGAVAVASTAGAPT 617
QY 615 SAPAPSSSA-SGPNQVICTRLVLSCPRALRLHYGONGGERPEKCKVCGRAFTGRNLRAH 673
DB 618 SAPAPSSSASGPNQVICTRLVLSCPRALRLHYGONGGERPEKCKVCGRAFTGRNLRAH 677
QY 674 FVGHKTSPPARAONSCPTCOCKKFTNAVTLQOHVNRHNLGQITNGSGALSSEGGAAQENSS 733
DB 678 FVGHKASPARAONSCPTCOCKKFTNAVTLQOHVNRHNLGQITNGSGALTALBEGGAADENG 737
QY 734 EOSTASGPGSPPOQSOOPSPSEEMS--EEDDEDEEDVDTEDESLAGRSESGEKAIS 792
DB 738 EOSTVSGAGSPPOQSOOPSPSEELSEEDDEDEEDVDTEDESLAGRSESGEKAIS 797
QY 793 VRGDSSEVSAGEEVATVVAAPTVEKMDSNKADONLPPPPPDNDLHPQPMEDGTS 852
DB 798 VRGDSSEVSAGEEVGVVAATAATAGKEMDSNEKTTQOSSLPPPPDSDLPQPMEDGSS 857
QY 853 DVSGAMEEELKLEGISPMALTOEGESTPTVLEELNPEAMKRDGESSGRKACVCG 912
DB 858 GYVGKKEGEGKPERSSSPASALTPEGEATSVLVEELSLQEAARKKEGESSSRKACEVCG 917
QY 913 QSPPTOTALPEHOKTHPPKDGFLTCVFCROGFLDRATLKHMHLAHQVPPFAPHGONI 972
DB 918 QAFPSQALAEHOKTHRKEGFLFTCVFCROGFLERATLKKHMLAHQVPPFAPHGONI 977
QY 973 ATLSLVPGCSSIPSPLSPFPRKDDPTMP 1002
DB 978 ALSTLVPGCSPSTSTGSLSPFPRKDDPTIP 1007
RESULT 7
AAB93193
ID AAB93193 standard; Protein; 813 AA.
XX
AC AAB93193;
DB 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12150.
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
OS Homo sapiens.
PN EP1074617-A2.
XX
XX 07-FEB-2001.
PD 28-JUL-2000; 2000EP-0116126.
PF

XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 12150; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo or primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 represent human amino acid sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 813 AA:

Query Match 15.9%; Score 845.5; DB 22; Length 813;
Best Local Similarity 30.5%; Pred. No. 1.4e-44;
Matches 242; Conservative 101; Mismatches 265; Indels 185; Gaps 27;

Y 253 ASSSSSSSSGAEPRQAFFHLVHPLG-----SQHPPSVGVGGRSHKPTPASPALPGS 306
b 3 ASHAIHSSGAGADTLKTLGSHMSQOVSAAVALLSQKAGSQGLSDALKQAKLPHANIPSA 62
Y 307 TDOLIASPHLA-----FPGTTGLAACLGAARGEAASP-----GLKPKNS 351
b 63 TSSL--SGCLAPFLTKPDCTRVLPVNWMSRLPSALLPOAPGVSLFQSPSTVALDTSKKKK 120
Y 352 GELGYGEVYISLE-KPGG-----RHKCRFCAKYVGSALQIHLRSHTERPYKNCVGN 405
b 121 GK---PWNISAVDVKPKDEALYKHKCKYCSKVFGTDSLIQIHLSRHGERPYVSCVGH 177
Y 406 RFTTRGNLKVHFHRHREKYPHVQOMPVPVPEHLDYVITSSGLPYGMSVPEPKAEAEAGTP 465
b 178 RFTTKGNLKVHFHHRH---POVKANPOLFAEFODKVAAGNGIYPALVSP-----DPIDEP 228
Y 466 GCGVERPVLVSTALSTAESLTILSTGTSTANAPGLPTNPKFVLMKAVEPKSKADEMTIP 525
b 229 SLSDSKRPVLV-TTSVGLPQN---LSSGT-----NPKDLTGSGSLPGDRPGPSP 273
Y 526 PGSEGSALAGVADSGSATRMQLSKLVTSILPSWALLTNHLKSTGSPFPYVLEPIGASPSSE 585
b 274 ESEBGQPLTPGVGPNVNSPR-----AGGFQSGSTPEP-----GSE 307

OY 586 TSKIQQLVEKIDROGAAVAVASTASGAPTTSDAPSSSSAGPNOCVLCIRVLSCPRALRLH 645
Db 308 TLKIQQLVENVNDK-----ATPDNPECLICHRVLSQSSSLKMH 344
OY 646 YGQHGGERPFCKCYKGRAPSTRGNLRAPFVGHKTPSAARQNSCPICOKKFTNAYVTLQOH 705
Db 345 YRTHTGRPFQCKICGRAPSTKGNLTKLGVHRTWNTSIKTQHSCTPICQKKFTNAYVTLQOH 404
OY 706 VRNHLGGQIPN-----GGSALSEGGGA-----OENSSEQSTAS 739
Db 405 IRMHMGQIDNTPLPENPCDFTGSEPMVTVGENGSGALICHDVDIESIDVEVSSQEARPS 464
OY 740 G---PGSFPPQSQPSPPEEEMSEBEDEDEEEDVTDEDSIAG-----RGS-ESGG 787
Db 465 SSRVPTPLPSIHSAFPLIGFAMMAS-----LDAPGRVGPAPFNLQGRSRENGS 513
OY 788 EKATSVRGDSEEVSGAEEVATSVAAPTTVKEMDSNEKAPQHTLPPPPPPDNLDHQPM 847
Db 514 VESDGLTNDSSSLMGDOEYQSR-----PDILETTSPQALSPANQOAESIKSKSP- 563
OY 848 EOGTSDVSGAMEEAEKLEGISS-----PMAALTOEGEGTST-PLVBEINLPEAMKRD 898
Db 564 -DAGSKAESSENSRTMEGRSSLPSTFIRAPPTYKVEPGTFVGBSTLSPGMTPLLAQ 622
OY 899 PGSSGRKACEVCGSPPTQTALEHOKTHRKDGPLFTVCFRQGLDRATYTKKMLLAH 958
Db 623 PRQAKQHGCTRGKKNSSASALQIHERHTGKERP-FVCNITCGRAFTTKGNLKVH----- 676
OY 959 HOVPPAPHPQPN 971
Db 677 -----YMTGANN 684

RESULT 8
ABB62368
ID ABB62368 standard; Protein; 1373 AA.
XX
AC ABB62368;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13896.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX N-PSDB; ABL06471.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PS
XX Disclosure; SEQ ID NO 13896; 21pp + Sequence Listing; English.
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is

useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB130511), expressed DNA sequences (AB101840-AB16175) and the encoded proteins (AB85737-AB12072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 1373 AA:

Query Match 14.4%; Score 765; DB 22; Length 1373;

Best Local Similarity 22.9%; Pred. No. 3.1e-39;

Matches 319; Conservative 151; Mismatches 446; Indels 476; Gaps 46;

3 QETSSSRIGPCGPAERNGDASEEHNPQVCAKCAQFSDPTEFLAHQNSCTDPVAV 62
50 KDIGSDQENGCGSPLTATTTASPSRSP-----PEEQPEEQSTSEOST---PEQST 100
63 IIGGQENPNSASSAPREGHSRQVMTEDHSN---PPDSGSGPPDP----- 109
101 PDHLENIKSEAKSEIEVEDNNRVAMTKPPSEEREPPNAGSMSPSVAAEASAEAT 160
110 -WGPRRGE-----ESSGQFLVAATGTAAAGGGGLI-----LASPKIGAT 148
161 ERTPEKEKEDVENVDEKDEARSSAVPSTEVTLPGGAGAVYTELAIQMMAIAQFAAK 220
149 PLPESTAPRPPPPPPGCVSGHNLPLLEELRYLQORQHOMQTEQICROVLL- 207
221 TIANGSNGADNE-----AAMKQLAFLOQTLFNLQOQQLFOIQLIQLOSLALN 269
208 -----LGSIGQTVGAPASPELPTGAASTRKPL 237
270 QAKQEDTEEDADQEQDQETDYEEERLADMLRQAKERMAE-----AKAQHLL 323
238 PLFSPKPAQTGKTATSSSSSSSSGAERPKQAFNLHPGLSGHPFSGVGRSHKPTP 297
324 NAGVPLR-----ESSGSPAESLKRHRREHDHESQPNRR-----SLDNTKADT 366
298 A-----PSPALPGSTD--OLIASPH--LAFPGTTGLLAQCLGAARGLAASPG 343
367 AODALAKLEKEMENTPLRFSGDLASSIITNHDLPRNSLDLQKR---AQEVLDSASOG 422
344 LK-----PKNSGELGVEGVESLSLEKPGRHKCRFCACVSGDSALQIHLRSH 393
423 ILANSADPAFEKESGEK-----GRNEPFFKHRCRYCGKVFSGDSALQIHRSHT 474
394 GERPKVCNVCGNFTTNGNLKVHFRHREKYPHYQMNPHVPEH-----LDVYTSS 445
475 GERPKVCNVCGSFTTNGNLKVHFRHREKYPHYQMNPHVPEH-----LDVYTSS 445
446 GLPYGMSVPRKEAEELGTEGG--GVER--KPLVASTALSTESTLLSTGSTAV--- 498
535 SSPNHSRPPRLSGAPSPAPRGLQMLYRPMELIKSLGAAPRHOYFQELPTDLRKP 594
499 AGLPTFNKFLVLAAPKSKADENTPRSGES-----AIAGVADSGSATRMQLSK 549
595 SPQLDEDEPVYKNEPVEKQRENEDEMAESSEPERPLREVKREKEVEDEVOYKQE 654
550 -----LVTSPLSWALLTNLKGTSFPFVYLERL-----GASPSFTSKLQOL-- 592
655 DHRLEPRRTSPSSSEHSRNNHNRSHMGYRVVYQPIQAPALMHPSSPGSQSHDHLPTP 714
593 -----VEKIDROGAVAVASTASGAPTTSAPAR----- 619
715 GOLPRPREDFAERPLNFTTAKMLSPREHNSPVRSRPGALRPGVPRPHNPHMASSPF 774
620 -----SSSASGPNQVCILRYLS 637
775 FNPDKHEMALLPKSHNSDNEMENFIEVSNTCETMKLKEIMKKKKISDPNQCVCADRYLS 834
638 CPRALRLHYOHGGERPRPKVCYCGRAFTRGMLRAHFVGHKITSPPARAQNSCPLCQKKFT 697

835 CKSALQNMVHRTHTGERPFCRIQGRAFTTKGNLKTNHAVNHKIRPRMRFHOCYCHKYS 894
698 NAYVIAQVHVMILG-----GQIPNGSALSEGGAQENS 733
895 NALVIAQVHVMILG-----GQIPNGSALSEGGAQENS 733
734 EOSTASGSPFPPOP-----OSQSPSEEMSEEEDE-----EEEE 770
949 ---LFGGPGPGPPRNGAHNHALGSESSQGDMDMOCGEDYDDDVSSPHLSNMLEQ 1005
771 D-----VTDESLAGRSGESGGEKA---ISVAGDSE 798
1006 DRSRGDDFKSLLEQKLRIDATGVVNTNPRRPSASHSVSTSAPTSPSVASSQ 1065
799 EV-----SGAEVFA-----TSVAAPTVMKMS-----NEKAPQH 829
1066 VIKRSSPAPSEASQALDILTPRAAPTSSSSSRSLPKEKVPSPSLRPSGSSHASAN 1125
830 TLPPPPP-----PDNLDP-----POPMQ 849
1126 ILTSLPLPTVYGLDCLPRGLQHNHQQQHNHLMQQAVALAAAHQNNHQAALHONQEQ 1185
850 GTSVSGAMEEBAKLEGISSPMAAL-----TQBG-----EQTSTPLVE 887
1186 LRREAEAOQKAAAAAAAAAAAAOQROTPOARDQEGGAGRPPLMGARPPGM 1245
888 EUNLP-----EAMK-----DPGESSGRK---ACEVGSQSPPTQALE 922
1246 FPLPLPPTATTONCMAMMOIAOSVMPARPNPLALSVCBSTTGCICKTTPCHSALE 1305
923 EHQKTHPKDPLFTYVFCRQGLDRLATLKMLLHNOVPRAPRPHQONTATLSLVGCS 982
1306 IYRSHTKERP-FKCSICDRGFTTKGNLKNHMLT--HKIRMQEQTFRNRAVKYV----- 1357
983 SSISPGLSPFP 994
1358 -SLP---ISLP 1365

RESULT 9
ABB61050
ID ABB61050 standard; Protein; 1267 AA.
XX
AC ABB61050;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 9942.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX MPI; 2001-656860/75.
XX
XX N-PSDB; ABL05153.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT

polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoietic regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

sq	Sequence	172 AA;
	Query Match	13.1%;
	Best Local Similarity	84.4%;
	Matches 135; Conservative	4; Mismatches 21; Indels 0; Gaps 0.

SQ Sequence 172 AA;

Query Match	13.1%	Score 697	DB 22	Length 172
Best Local Similarity	84.4%	Pred. No. 4,1e-36		
Matches 135, Conservative	4	Mismatches 21	Indels 0	Gaps 0

80 RPEGHSKRSQVMDIETHSNPDDSGSSGPPDPITWGEERKEEESGGLVNAIGITAAAGGGGL 13
 11 RREVSNNQVRKKEHSIPDDYSSVPTDPTWGERREEESSGHFMVDHTGTAAAGGGGMI 70

71 LASPLGATPLPEXAPAPPPPPPPGVSGHLNPLLEELRVLQQRQIHQMOMTE 130

Db 131 QICROVLLGSLGQTVGAPASPSEIHGTASTKPLPL 170

XX	ID	standard; Protein; 200 AA.
	AAB92796	

XX 26-JUN-2001 (first entry)
DT
XX
Human Proteome Sequence Set ID NO:11208

XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy
KW	
XX	Homo sapiens.
OS	

PN EP1074617-A2.
XX
PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR

PR 09-JUN-2000; 2000JP-0241899.
XX

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Salto K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX

XX	primer sets for synthesizing polynucleotides, particularly the 5602
PT	full-length cDNAs defined in the specification, and for the detection
DR	WPI; 2001-318/49/34.

PT	full-length cDNAs -
XX	
PS	Claim 8; SEQ ID 11298; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 502
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementa

to the complementary strand of apolynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 3'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily by using specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AA82246 to AA895893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

SQ	Sequence	200 AA;		
Query Match		10.0%;	Score 529;	DB 22;
Best Local Similarity		82.2%;	Pred. No. 1.6e-25;	Length 200;

[illegible]

Qy SSAPRQEGHSRQVMDTEHNSPPDQSGSSGPPDDPTWGPERRQEESGQFLVAATGTAAG 133

76 SSAPRQEGHSRQVMDTEHNSPPDQSGSSGPPDDPTWGPERRQEESGQFLVAATGTAAG 133

Dh SSAPRQEGHSRQVMDTEHNSPPDQSGSSGPPDDPTWGPERRQEESGQFLVAATGTAAG 135

RESULT 12
ABB30894

XX
AC ABB30894;
XX
XX 01 Jun 2003 (first entry)

XX	peptide #3545 encoded by breast cell single exon nucleic acid probe.
DE	
XX	
XX	Human: microarray; single exon probe; gene expression; breast;

xx
xx
OS
Homo sapiens.
xx

XX
PD
XX
09 - AUG - 2001.
XX
30 - JAN - 2001 - 2001WC-US000662.
PE

```

AA      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.

```

PR 21-SEP-2000: 2000US-0234687.
PR 27-SEP-2000: 2000US-0236359.
PR 04-OCT-2000: 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX

XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT useful for measuring gene expression in sample derived from human

[illegible]

Db	270	ESQSAGSPAISESTSSQALSPSNSTQEFHKS-----PSIEEKQRAVPSEFANG	319
Oy	835	- PPPPDN 840	
Db	320	LSPTPVN 326	
RESULT 15			
AAM69245			
ID	AAM69245	standard; Protein; 330 AA.	
XX	AC	AAM69245;	
XX	DT	06-NOV-2001 (first entry)	
XX	DE	Human bone marrow expressed probe encoded protein SEQ ID NO: 29551.	
XX	XX	Human; bone marrow expressed exon: gene expression analysis; probe:	
KW	microarray; cancer; leukemia; lymphoma; myeloma.		
XX	OS	Homo sapiens.	
XX	PN	MO200157276-A2.	
PD	09-AUG-2001.		
XX	30-JAN-2001; 2001WO-US00668.		
XX	PR	04-FEB-2000; 2000US-0180312.	
PR	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
PR	04-OCT-2000; 2000GB-0024263.		
XX	(MOLE-) MOLECULAR DYNAMICS INC.		
XX	PI	Penn SG, Hanzel DK, Chen W, Rank DR;	
XX	DR	WPI; 2001-488900/53.	
XX	PT	Human genome-derived single exon nucleic acid probes useful for	
XX	PT	analyzing gene expression in human bone marrow -	
XX	PS	Example 4; SEQ ID NO: 29551; 658bp + Sequence Listing; English.	
XX	CC	The present invention provides a number of single exon nucleic acid	
CC	CC	probes which are derived from genomic sequences expressed in the human	
CC	CC	bone marrow. They can be used to measure gene expression in bone marrow	
CC	CC	samples, which may enable the improved diagnosis and treatment of cancers	
CC	CC	such as lymphoma, leukemia and myeloma. The present sequence is a	
XX	XX	protein encoded by one of the probes of the invention.	
XX	XX	Sequence 330 AA;	
XX	XX	9.1%; Score 482.5; DB 22; Length 330;	
XX	XX	Best Local Similarity 35.7%; Pred. No. 2.5e-22;	
XX	XX	Matches 131; Conservative 37; Mismatches 98; Indels 101; Gaps 13	
Oy	531	SAIVGADVDSGNSATRMQSLKTVTLSPWMLLTNNLKTSGSEPPFVLEPLGASPSSETSKIQ 590	
Db	4	SPADDCGPAAGSATTE-----TNPLPLPMSDQFKA--KPPFGGLD--SAQASSETSKIQ 52	
Oy	591	QLVEKIDRQAGAVAVASTASGAPTTSPAPBSSASGPNQCVLCPLVLSQPRALRLHYGONG 650	
Db	53	QLVLENIDRK-----ATDPNECITICHRVLSQSAALKMHTRTHT 89	
Oy	651	GERPFKCYVCGAFSTRCNLRHFVGHKTSPPARAQNSCPICQKFTNAVTLQGHVRL 710	
Db	90	GERPFKCYICGAFSTRCNLRKTHYSVHRAPRLRYGHSQPCIQKFTNAVTLQGHVRL 149	
Oy	711	GGQIPN-----GGSALSEGGAQDEN-----SSQSTASGPGSPQ-QPSQSPSE 755	

```

|||||
150 GGOIPNTVPYDSTYSESMSDPTGSEFDEKFNFDLDNFSDENMEDCPGSIPTPKSADASQD 209
756 E-----EMSE-----EEEEDEEEEDVTDEDSTLAGRG-- 782
210 SLSSSPFLPLEMSSIALLENQMKMINAGLAEDLOASLKSVENGSIEGDVLITNDSSSVGGDM 269
783 -SESGGEKAISVRGDSSEV---SGAEEVATSVAAPTTVKEMDSNEKAPQHTLPP----- 834
b 270 ESQASGSPAISESTSSMQALSPNSTQDEFHKS-----PSIEKPKQRAVSEPFANG 319
y 835 -PPPPDN 840
b 320 LSPTPVN 326

ESULT 16
AA17074
AA17074 standard; Protein; 330 AA.
AA17074;
12-OCT-2001 (first entry)
Peptide #3508 encoded by probe for measuring cervical gene expression.
Probe; human; microarray; gene expression; cervical epithelial cell;
cervical cancer.
Homo sapiens.
W0200157278-A2.
09-AUG-2001.
30-JAN-2001; 2001WO-US00670.
04-FEB-2000; 2000US-0180312.
26-MAY-2000; 2000US-0207456.
30-JUN-2000; 2000US-0608408.
03-AUG-2000; 2000US-0632366.
21-SEP-2000; 2000US-0234687.
27-SEP-2000; 2000US-0236359.
04-OCT-2000; 2000GB-0024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-488901/53.
Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human cervical epithelial cells -
Claim 27; SEQ ID No 21900; 487pp; English.
The present invention relates to human single exon nucleic acid probes
(SENP: see A110068-A118459). The present sequence is a peptide encoded
by one such probe. The SENPs are derived from human HeLa cells. The SENPs
can be used to produce a single exon microarray, which can be used for
measuring human gene expression in a sample derived from human cervical
epithelial cells. By measuring gene expression, the probes are therefore
useful in grading and/or staging of diseases of the cervix, notably
cervical cancer.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
Sequence 330 AA;
Query Match 9.1%; Score 482.5; DB 22; Length 330;
Best Local Similarity 35.7%; Pred. No. 2,5e-22;
Matches 111; Conservative 37; Mismatches 96; Indels 101; Gaps 13;
```

```

QY 531 SAIAVADSGSATRMQSLKLVTSIPSWALLTNHLKSTGSEFPFVYLEPLCASPSETSKLQ 590
| | | | |
Db 4 SPAAOCGPAGSATTF-----TNPLLPLMSEQFKA--KFPFGGLID--SAQASETSKLQ 52
QY 591 QLVKIDRGAAVAVASTASGAPTTASAPASSASGPNOCVICTLRVLSCPRALRLHYGHG 650
| | | | |
Db 53 QLVEMIDKR-----ADDPNECIIHRVLSCOSALMHTYHT 89
QY 651 GERPFKCKVCGRAFTSRGNLRAHFVGHKTSAPAARQNSCPTQCKRTNAVTLQOHVAMHL 710
| | | | |
Db 90 GERPFKCKICGRAFTTKGNLKHYSVHRAMPRLRVQHSICPCIQCKKFTNAVVLQOHTRMM 149
QY 711 GGOIPN-----GGALSISGGGAQEN-----SSFOSTASGSGSFQ--POSQPSPE 755
| | | | |
Db 150 GGOIPNTVPYDSTYSESMSDPTGSEFDEKFNFDLDNFSDENMEDCPGSIPTPKSADASQD 209
756 E-----EMSE-----EEEEDEEEEDVTDEDSTLAGRG-- 782
Db 210 SLSSSPFLPLEMSSIALLENQMKMINAGLAEDLOASLKSVENGSIEGDVLITNDSSSVGGDM 269
QY 783 -SESGGEKAISVRGDSSEV---SGAEEVATSVAAPTTVKEMDSNEKAPQHTLPP----- 834
| | | | |
Db 270 ESQASGSPAISESTSSMQALSPNSTQDEFHKS-----PSIEKPKQRAVSEPFANG 319
QY 835 -PPPPDN 840
Db 320 LSPTPVN 326

RESULT 17
AA17074
AA17074 standard; Protein; 330 AA.
AA17074;
17-OCT-2001 (first entry)
Peptide #3605 encoded by probe for measuring placental gene expression.
Probe; microarray; human; placenta; antenatal diagnosis;
genetic disorder.
Homo sapiens.
W0200157272-A2.
09-AUG-2001.
30-JAN-2001; 2001WO-US00663.
04-FEB-2000; 2000US-0180312.
26-MAY-2000; 2000US-0207456.
30-JUN-2000; 2000US-0608408.
03-AUG-2000; 2000US-0632366.
21-SEP-2000; 2000US-0234687.
27-SEP-2000; 2000US-0236359.
04-OCT-2000; 2000GB-0024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-48897/53.
Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human placenta -
Claim 27; SEQ ID No 29837; 654pp; English.
The present invention relates to single exon nucleic acid probes (SENP:
see A11315-A115746). The present sequence is a peptide encoded by one
such probe. The probes are useful for producing a microarray for
```


hyaline membrane disease.
Homo sapiens.
WO200186003-A2.
15-NOV-2001.
30-JAN-2001; 2001WO-US00665.
04-FEB-2000; 2000US-180312P.
26-MAY-2000; 2000US-207456P.
30-JUN-2000; 2000US-0608408.
03-AUG-2000; 2000US-0632366.
21-SEP-2000; 2000US-234687P.
27-SEP-2000; 2000US-236359P.
04-OCT-2000; 2000GB-0024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2002-114183/15.
Spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples -
Claim 27; SEQ ID NO 28520; 634pp; English.
The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array. Identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Best Local Similarity 35.7%; Pred. No. 2.5e-22;
Matches 131; Conservative 37; Mismatches 98; Indels 101; Gaps 13;
QY 531 SAAGVADSGSATRMOLSKLVTSLPSWALLTNHLKSTSPFPYVLEPLGASPSSTKIQ 590
| : ||||| : : : : : ||| : : : |||||
Db 4 SPNADCGPAGSATTF-----TNPLPLMSEQFKA--KFPFGGLD--SAQASSTSKIQ 52
QY 591 QLVKIDRQGAVAVASTASGAPTTASAPSSASGPNOCVICILRLSCPRLALHGOHG 650
||| ||||| : : ||| ||||| |||||
Db 53 QLVENIDKK-----ATDPNCCIIICHRVLSOSALKMHRTH 89
QY 651 GERPEKCVGGAFTSTRGNLRAHFVGHKTSAPARAQNSCPICQKKFTNAVTLQOHRMHL 710
||| ||||| : : ||| : : ||| ||||| |||||
Db 90 GERPEKCKICGRAFTTKGNLKHYSVHRAMPPLRVGHSCPTCQKKFTNAVTLQOHRMHL 149
QY 711 GGOJPN-----GGALSREGCAQDEN-----SSQSTASGSGSPFQ--PQSQPSPE 755
||| ||||| : : : : : ||| : : ||| : :
Db 150 GGOJPNTPVPDYSSEMSDSTGSEDEKNFDLDNFSDEMDCPGSGIDTPKSAVASQD 209
QY 756 E-----EMSE-----EEEEDEEEDVYDEDSLARG-- 782
Db 210 SLSSPFLPLEKSIATLENQMKMINAGLAEQLQASLKSEVENSITGCDVLTNDSSVYGDM 269
QY 783 -SESGEKATSVRGDSEEV---SGAEEVATSVAPPTVKEMDSNEKADQHTLPP--- 834
| : | ||| : : : : : | : : |||
Db 270 ESQASGSPATSESTSMQALSPNSTQGFHKS-----PSIEEKQRAVPSEFANG 319
QY 835 -PPPPDN 840
| : |
Db 320 LSPPEVN 326
RESULT 20
ID ABB32272 standard; Peptide; 336 AA.
XX ABB32272:
AC XX
XX 01-FEB-2002 (first entry)
DT XX
XX Peptide #4923 encoded by breast cell single exon nucleic acid probe.
DE XX
XX Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
OS Homo sapiens.
XX
XX WO200157271-A2.
PN 09-AUG-2001.
PD XX
PF 30-JAN-2001; 2001WO-US00662.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
DR
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 15240; 327pp + sequence listing; English.
XX

CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 336 AA:

Query Match 8.3%; Score 438.5; DB 22; Length 336;
Best Local Similarity 30.4%; Pred. No. 1.5e-19;

Matches 128; Conservative 42; Mismatches 128; Indels 123; Gaps 13;

QY 510 LMKAVEPKSKADENTP-----PGESEGS-AIAGVADSGSATRMQLSKLVTSLPSMALL 560
DB 5 LSSGTPNPDLTGSLPGDLQPPSPSESEGGPTLPQVGPVNSPR----- 48
QY 561 TNLHKTSGSFPPRYVLEPLGASPSFETSKLOOLVERKIDRQGANAVASTAGCAPTTSAPAPS 620
DB 49 -----AGGFQSGSTPEP-----GSETLKLQDLVENIDK----- 76
QY 621 SSASGPNOCVLCRLVLSCPRALRLHYGOHGERPFCKVCYGRAFSRGNLRAHFVGHKTS 680
DB 77 -ATTDPNECLICHVLSQSSLSKMHYRTHTGERPFQCKICGRAFSRKNLKTHLGVHRTN 135
QY 681 PAARAQNSCPICQKKTNNATLQOHVNRHMLGGQIPNGSALSSEGGAAQENSSSQSTASG 740
DB 136 TSIKTOHSCPTICQKKTNNATLQOHVNRHMLGGQIPN----- 171
QY 741 PGSFPOQSQPSPPEEEMSEEEDEEDVTDEDSLARGSESGEKA1-----SV 793
DB 172 -----TLPENPC-----DFTGSEPMT--VGENGSTGAICHDDVYESI 207
QY 794 RGDSEEVSGAEEEVATSVAAPTTVKEMDSNEK---APOHTLPP--PPPDNLDPQP 846
DB 208 --DVEEVS--SQEAPSSSSKVPPLPSIHASPTLGFAMMASLDAFGKVGPAFPNLQROGS 264
QY 847 MEOGTSDVSGAMEEAKLEGISSPMAALTQEGEGTSTPLVEELNLPAMKKDPRESSGRK 906
DB 265 RENGSVESDGLTNDSSSLMGDOEQYQSRSPDILETTTSFOALSPANQAESTIKSPDAGSK 324
QY 907 A 907
DB 325 A 325

RESULT 21
ABB37530 ID ABB37530 standard; Peptide; 336 AA.
AC ABB37530;
XX 04-FEB-2002 (first entry)
XX Peptide #5036 encoded by human foetal liver single exon probe.
XX DE Human; foetal liver; gene expression; single exon nucleic acid probe.
XX KW Homo sapiens.
XX OS

PN W0200157277-A2.
XX 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US00669.
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -

PS Claim 27; SEQ ID NO 30165; 639pp + sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 336 AA:

Query Match 8.3%; Score 438.5; DB 22; Length 336;
Best Local Similarity 30.4%; Pred. No. 1.5e-19;

Matches 128; Conservative 42; Mismatches 128; Indels 123; Gaps 13;

QY 510 LMKAVEPKSKADENTP-----PGESEGS-AIAGVADSGSATRMQLSKLVTSLPSMALL 560
DB 5 LSSGTPNPDLTGSLPGDLQPPSPSESEGGPTLPQVGPVNSPR----- 48
QY 561 TNLHKTSGSFPPRYVLEPLGASPSFETSKLOOLVERKIDRQGANAVASTAGCAPTTSAPAPS 620
DB 49 -----AGGFQSGSTPEP-----GSETLKLQDLVENIDK----- 76
QY 621 SSASGPNOCVLCRLVLSCPRALRLHYGOHGERPFCKVCYGRAFSRGNLRAHFVGHKTS 680
DB 77 -ATTDPNECLICHVLSQSSLSKMHYRTHTGERPFQCKICGRAFSRKNLKTHLGVHRTN 135
QY 681 PAARAQNSCPICQKKTNNATLQOHVNRHMLGGQIPNGSALSSEGGAAQENSSSQSTASG 740
DB 136 TSIKTOHSCPTICQKKTNNATLQOHVNRHMLGGQIPN----- 171
QY 741 PGSFPOQSQPSPPEEEMSEEEDEEDVTDEDSLARGSESGEKA1-----SV 793
DB 172 -----TLPENPC-----DFTGSEPMT--VGENGSTGAICHDDVYESI 207
QY 794 RGDSEEVSGAEEEVATSVAAPTTVKEMDSNEK---APOHTLPP--PPPDNLDPQP 846
DB 208 --DVEEVS--SQEAPSSSSKVPPLPSIHASPTLGFAMMASLDAFGKVGPAFPNLQROGS 264
QY 847 MEOGTSDVSGAMEEAKLEGISSPMAALTQEGEGTSTPLVEELNLPAMKKDPRESSGRK 906
DB 265 RENGSVESDGLTNDSSSLMGDOEQYQSRSPDILETTTSFOALSPANQAESTIKSPDAGSK 324
QY 907 A 907
DB 325 A 325

RESULT 22
AA58186
ID AA58186 standard; Protein; 336 AA.
C
X
X
AA58186;
T
05-NOV-2001 (first entry)
X
X
Human brain expressed single exon probe encoded protein SEQ ID NO: 30291.
X
W Human; brain expressed exon; gene expression analysis; probe;
M microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
K epilepsy; cancer.
X
Homo sapiens.
F
WC0200157275-A2.
D
09-AUG-2001.
X
X
30-JAN-2001; 2001MO-US00667.
F
04-FEB-2000; 2000US-0180312.
R
26-MAY-2000; 2000US-0207456.
R
30-JUN-2000; 2000US-0608408.
R
03-AUG-2000; 2000US-0632366.
R
21-SEP-2000; 2000US-0234687.
R
27-SEP-2000; 2000US-0236359.
R
04-OCT-2000; 2000GB-0024263.
X
(MOL-) MOLECULAR DYNAMICS INC.
A
Penn SG, Hanzel DK, Chen W, Rank DR;
X
WPI; 2001-483446/52.
R
Single exon nucleic acid probes for analyzing gene expression in human
T
brains -
X
Example 4; SEQ ID NO: 30291; 650pp + Sequence Listing; English.
S
The present invention provides a number of single exon nucleic acid
C
C probes which are derived from genomic sequences expressed in the human
C brain. They can be used to measure gene expression in brain cell samples,
C which may enable the diagnosis and improved treatment of nervous system
C diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
C epilepsy and cancers. The present sequence is a protein encoded by one of
C the probes of the invention.
X
Sequence 336 AA;
Q
Query Match 8.3%; Score 438.5; DB 22; Length 336;
Best Local Similarity 30.4%; Pred. No. 1.5e-19;
Matches 128; Conservative 42; Mismatches 128; Indels 123; Gaps 13
Y
510 LMKAVEPKSKADETTP-----PGSEGS-AIAGVADSGSATRMQLSKLYTSLPWMALL 560
I : : :
5 LSSTGNKRDLGSLRFDLGRPRSPESGPTLRGVGYNINSPR----- 48
Y
561 TNNLKSTGPSFPFVYLEBGLASPSETSLSLOOLEKIDROGAVAVASASTAGATTSAPAFS 620
I
49 -----AGGFYGSGSTPREP-----GSETTKILOLVENIDK----- 76
Y
621 SSASGRPOCYLTALVLSGRRLRYHYOHNGBERRPKKCVCVCSRAFTSRGTLRHAFGHKTS 680
: : :: : :: : : : :
77 -ATTDPRECLICHNVULSCOSSLMKHMYRTNHTGERPROCKICGAFAETKNLKTITLVHRTN 135
Y
681 PAARAONSCPTCQKKFFNAVLTLOOHVNHLGGQILRPGSSALSSEGGAOENSDSTASG 740
: :
b 136 TSAIKTONSCPTCKKKFTNAVLDOHIRNMHGQIPN----- 171
Y
741 PGSPFOPOSOPPFDEEMSEEDDEDVEDVDSDSLAGNSGESGEKA1-----SV 793

[illegible]

```

RESULT 23
AAM70641
ID AAM70641 standard; Protein; 336 AA.
AC
XX
XX AAM70641;
DT 06-NOV-2001 (first entry)
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30947.
KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; Leukemia; Lymphoma; myeloma.
XX Homo sapiens.
OS
PN W0200157276-A2.
PD
XX 09-AUG-2001.
PF
XX 30-JAN-2001; 2001WO-US00668.
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488900/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
PS Example 4; SEQ ID NO: 30947; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
XX Sequence 336 AA:
SQ
Query Match 8.3%; Score 438.5; DB 22; Length 336;
Best Local Similarity 30.4%; Pred. No. 1.5e-19;
Matches 128; Conservative 42; Mismatches 128; Indels 123; Gaps 13;
OY 510 LMKAVPEPKRADENP-----PGSEGS-AIAGVADSGSATRKMLSKLYTSLPSMALL 560
DB 5 LSSGTPKRLTGGLGGDILPGPSPESSEGPTLPDPVPNTSPR----- 48

```


QY 561 TNHLKSGSFPPFYVLEPLGASPSSETSKLOQVLEKIDROGAVAVASTAGAPTTSA PAPs 620
CC 49 -----AGGFGSGTPEP-----GSETLKLQOLVENDK----- 76
CC 621 SSASGPNOCVICTRLVLSRPRALRLHYGOHGERPERPKVCYCGAFSTRGNLRAHFVGHKTS 680
CC 77 -ATTDPNECLICHRVLSGSSSLKMHYRTHGTGERPQCKICGRASTKGNLKTHLGVHRTN 135
QY 681 PARAO NSCPTCOCKKFTNATVLTQOHVHMLGGQIPNGSALSSEGGAAQENSSEOSTASG 740
CC 136 TSITQHSCTPCO KKFNTA VMLQOHIRMHMGQIPN----- 171
QY 741 PGSEFPQPOSQSPSEPEEMSEEEDEEEDVDDEDLSLGRSGESGGEKAI-----SV 793
Db 172 -----TPLPENPC-----DFTGSEPMPT--VGENGSTGALICHDDVIESI 207
QY 794 RGDSEEVSGAEBEVATSVAPTIVKEMDSNEK----APQHTLPP--PPPNLDHPOR 846
Db 208 --DVEEYS-SQEA P SSSSKVPTPLPSIHSA SP T L G F A M M A S L D A P K V G P A P F N L Q R G S 264
QY 847 MEGQTSDVSGAMEEAKLEGISSPMALTOEGEGSTPLVEELNLPEAMKKDPGESSGRK 906
Db 265 RENGSEVSDGLTNDSSSLMGDDQYQSRSPDILETTSTFOALSPANSQAESIKSKSPDAGSK 324
QY 907 A 907
Db 325 A 325

RESULT 24
AAM18485
ID AAM18485 standard; Protein; 336 AA.
AC AAM18485;
DT 12-OCT-2001 (first entry)
DE Peptide #4919 encoded by probe for measuring cervical gene expression.
KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer.
XX Homo sapiens.
XX OS
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX PD
XX 30-JAN-2001; 2001WO-US00670.
XX PF
XX 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX DR
XX Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 27; SEQ ID NO 2311; 487bp; English.

CC The present invention relates to human single exon nucleic acid probes
CC (SENPs: see A110066-A128459). The present sequence is a peptide encoded
CC by one such probe. The SENSEs are derived from human HeLa cells. The SENSEs
CC can be used to produce a single exon microarray, which can be used for

CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 336 AA:
Query Match 8.3%; Score 438.5; DB 22; Length 336;
Best Local Similarity 30.4%; Pred. No. 1.5e-19;
Matches 128; Conservative 42; Mismatches 128; Indels 123; Gaps 13;
QY 510 LMKAYEPKSKADNTP-----PGSEGS-AINGVADSGASTRMQSLVTSLSPEMALL 560
Db 5 LSSGTPKDLTGSLPGDLOPGSPSESEGGPTLPGVGNVNSPR----- 48
QY 561 TNHLKSGSFPPFYVLEPLGASPSSETSKLOQVLEKIDROGAVAVASTAGAPTTSA PAPs 620
Db 49 -----AGGFGSGTPEP-----GSETLKLQOLVENDK----- 76
QY 621 SSASGPNOCVICTRLVLSRPRALRLHYGOHGERPERPKVCYCGAFSTRGNLRAHFVGHKTS 680
Db 77 -ATTDPNECLICHRVLSGSSSLKMHYRTHGTGERPQCKICGRAFASTKGNLKTHLGVHRTN 135
QY 681 PARAO NSCPTCOCKKFTNATVLTQOHVHMLGGQIPNGSALSSEGGAAQENSSEOSTASG 740
Db 136 TSITQHSCTPCO KKFNTA VMLQOHIRMHMGQIPN----- 171
QY 741 PGSEFPQPOSQSPSEPEEMSEEEDEEEDVDDEDLSLGRSGESGGEKAI-----SV 793
Db 172 -----TPLPENPC-----DFTGSEPMPT--VGENGSTGALICHDDVIESI 207
QY 794 RGDSEEVSGAEBEVATSVAPTIVKEMDSNEK----APQHTLPP--PPPNLDHPOR 846
Db 208 --DVEEYS-SQEA P SSSSKVPTPLPSIHSA SP T L G F A M M A S L D A P K V G P A P F N L Q R G S 264
QY 847 MEGQTSDVSGAMEEAKLEGISSPMALTOEGEGSTPLVEELNLPEAMKKDPGESSGRK 906
Db 265 RENGSEVSDGLTNDSSSLMGDDQYQSRSPDILETTSTFOALSPANSQAESIKSKSPDAGSK 324
QY 907 A 907
Db 325 A 325

RESULT 25
AAM30956
ID AAM30956 standard; Protein; 336 AA.
AC AAM30956;
DT 17-OCT-2001 (first entry)
DE Peptide #4993 encoded by probe for measuring placental gene expression.
KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX KW
XX Homo sapiens.
XX OS
XX PN WO200157272-A2.
XX PD
XX 09-AUG-2001.
XX PF
XX 30-JAN-2001; 2001WO-US00663.
XX PR
XX 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.

Db 265 RENGSVESDGLTNDSSILMGDOEYGRSPDILETTTSQLSPANSQAESIKSPDAGSK 324
QY 907 A 907
Db 325 A 325

RESULT 27
ABG40328 standard; Peptide: 336 AA.
XX
AC ABG40328;
DT 19-AUG-2002 (first entry)
DE Human peptide encoded by genome-derived single exon probe SEQ ID 29993.
XX
KW Human: single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenen syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
PN WO200186003-A2.
PD 15-NOV-2001.
PE 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632386.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
PS Claim 27; SEQ ID NO 29993; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements for the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray, assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several

CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a peptide/protein
CC encoded by a single exon probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 336 AA:
QY 510 LMKAVEPKSKADENTP-----PGSEGS-AIACVADSGSATRMQLSKVTSLPSMALL 560
Db 5 LSSGTNPKDLTGSLPGDLOGPSPSESGPPLPGVGPYNSPR----- 48
QY 561 TNLKSTGSPFPYVLEPLGASPSSTSKLOQLVEKIDROGAVAVASTAGAPTTSAAPAS 620
Db 49 -----AGFGGSGTPEP-----GSETLKQQLVENIDK----- 76
QY 621 SSASGPNQCVICLRYLSCPRALRLHYGGRGEPFKCYCGRAFSTRGNLRAHFVGKTS 680
Db 77 -ATDPNKECLICHRVLSGSSILKMHYRTHTGERPFCKICGRAFSPTGNTKTHLGVRHTN 135
QY 681 PAARAQNSCPTCKOKKFTNAVTLQOHVBMHGGIIPNGSALSBSGGAAQENSBSQSTASG 740
Db 136 TSITQHSCTPCCKKFTNAVMVLQOHIRMHMGQIIPN----- 171
QY 741 PGSEFPQSQGPSPSEEMSEEEDEEDVDYDEDSLARGSESSEGEKAI-----SV 793
Db 172 -----TPLEPNPC-----DFTGSEPT--VGENGSTGAICHDDVIESI 207
QY 794 RGDSEENVSGAEDEVATVAAPTTVKEMDSNEK-----APQHTLPPP-----PPPDNDHPQP 846
Db 208 --DVEEVS--SSEAPSSSSKVPPTPLPSIHSAAPLIGFAMMASLDAIPKVGVPAPFNLRQGS 264
QY 847 MEGGTSDVSGAMEEAKLEGISSPMALNQGEGCTSPPLVEELNLPEAMKKDPGSSGRK 906
Db 265 RENGSVESDGLTNDSSILMGDOEYGRSPDILETTTSQLSPANSQAESIKSPDAGSK 324
QY 907 A 907
Db 325 A 325

RESULT 28
AAM80283
ID AAM80283 standard; Protein: 927 AA.
XX
AC AAM80283;
DT 06-NOV-2001 (first entry)
DE Human protein SEQ ID NO 3929.
XX
KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

Query Match	Best Local Similarity	Score	DB	Length	gaps
Matches 227; Conservative 92; Mismatches 402; Indels 350; Gaps 45;					
Sequence 927 AA;					
6.48; Score 337; DB 22; Length 927;					
21.2%; Pred. No. 1.2e-12;					
PGC-HRSQVMDTEHSNPPDSSG--PPDPTWGE 113					
OEPMNSNSASSAPR-----PEG-HRSQVMDTEHSNPPDSSG--PPDPTWGE 113					
10 EKDPRIVGAQSVPRGRALKGLSPGLDLSAFILFPPPRAGPWNITAVLSSGMDPTALMGPD 69					
RGEESSGGQFLVAATGTAA-----GGGGGLILASP-----KLGATPLPEESTPAP 158					
LOGPEQSPN--DARGAESENEEESPRESSGEETIMGDPQAPSPESKOSTEMSLERRSSGD 127					
114 RGEESGGQFLVAATGTAA-----GGGGGLILASP-----KLGATPLPEESTPAP 158					
128 PVPFONDPTP--LGHSN-PL-----DHQIPLD-----P 152					
159 PPPPPPPPPGVSGCHNIPILIEELANLQROQHOMQMTEDICHOVILLGSLQGVAP 218					
121 ASPSELTP-----GTGAASSTKPLPLPLFSP--IKPAQGTGTASSSSSSSS 262					
153 PAREVAVPTPSDWTKACASMQMGALTTTMSNPFPVPANPSRLRELVOGR-----200					
263 GAEPKQAEFHLYHLPLGSDHPFSGVGVGRSHKPTPADSPALPGSTDQILASPHLAFPGTT 322					

[illegible]

PF 28-JUL-2000; 2000EP-0116126.
PX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI: 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
PS Claim 8; SEQ ID 17076; 2537PP + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-qt primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 619 AA:
Query Match 6.2%; Score 330.5; DB 22; Length 619;
Best Local Similarity 21.4%; Pred. No. 1.8e-12;
Matches 173; Conservative 78; Mismatches 276; Indels 281; Gaps 34;
QY 218 PASPSSELPGTGAASSTKPLPLSPFIKAQOTGKTATSSSSSSSSGAE-----PKQ 269
DB 46 PADCSF-----EVAEVKP-----KPETEAKAEASGCEKVSAAKPPRYACPLCPKA 92
QY 270 AFHLVLPPLGSHQHFVSQGVGRSH---KTPAPSPALPGSTDLIAP-----HLAFPGT 321
DB 93 ---YKTAPELRSH-----GRSHTEGKPPPCP-----ECGRFEMPVCLRHVLA----- 132
QY 322 TGLLAOACIGAAAGLEAASAPGLRKPKNSSGELGVEVISLEKPGGRHKRCFCAKVEGS 381
DB 133 -----SHAGELPF-----RCAHCPKAYGA 151
QY 382 DSAQLIHLKRSHTGERPYPCNVCNFRFTTGNLKVHRRHREKRYHVOMNPHRVERHLDYV 441
DB 152 LSKLKIHQRGHTGERPYACADCGKSPADS---VFRKHRR--THAGLRP-----YS 197
QY 442 ITSSGLPAGMSVPPKAEAEAGTGGGVERRKPLVAT--TALSTAESLTLLSTGTSTAVA 499
DB 198 CERGGKAYA-ELKDLRNRHSHT-----GERPFLSCGSGSFSRSSSLTCHQRITAAQKP 251
QY 500 PGLPTFNK-FVLMKAVEPKSADENTPPGSEGSALAGVADSGSATRMQLSKLVTSLPMSWA 558

DB 252 YRCPACGKGFTQLSSVQSHERTHSGEKPELCPRCGRMFSDPSSFRRHQRAR----- 302
QY 559 LITNLKSTGSPFPFVLEPLG---ASPSETSKLOQLYEKIDRQGVANASTSGAPTTS 615
DB 303 ---EGVK-----PYACEKCGKDFROPADLA-MHRVHTGDR----- 334
QY 616 AAPASSASGPNOCVICLRVLSCPRALRLHYGOHGERPFKCKVCGRASSTGNLRAHFV 675
DB 335 -----PFKLCQDKFTFVASMDELKRHALVHSGQRFRCDEGGRATFAERASLTFHSR 384
QY 676 GHKTSAPAARAONSCTPIQCKFTNAVTLQOHVRLGQ---IPNG-----GSLSBEGGA 727
DB 385 VH-----SGERPFHCNACGKSPVSVSSLRKHERTHRSSEAGVPPAOELVYGLALPV--GV 438
QY 728 AGENSSEQSTASGPGSFPPQSQQSPREEKSEEEDEEEDVDVDEDSLAGRSGSG 787
DB 439 AGSSSAPAPAGAGIGDPPA-----GLLGLPPSSGG 468
QY 788 EKALSVRGDSEEVSGAEAEVATSVNAPTVKEMDSNEKAPQHTLPPPPPPDNLDPQM 847
DB 469 -----VMATQMQVYGM-----TVEHVECO 487
QY 848 EQGTSVSGAMEEBAKLEGISSPMALTOEGEGTSTPLYBELNLPAMKADPSSGSKA 907
DB 488 DAGVREAPGP-----EGAG-----EAGGEADEKPPQFV 517
QY 908 CEVCGGSPFPTQALEHOKTKPKDGLPTGVCFCRQGLDRLATLKKHMLAHNOVPF-AP 966
DB 518 CRCKETFEFTTLRLRRHRSHELRP-PPCTCGKGISDRAGLRKHS-RTHSSVPYICP 575
QY 967 HGFONITLSLVPGCSSSIPSGLSPPF 994
DB 576 HCKRAFLSADLRKHERRHPVMGTPTP 603
RESULT 30
AAM79299
ID AAM79299 standard; Protein; 869 AA.
XX
AC AAM79299;
XX
DT 06-NOV-2001 (first entry)
DE
DE Human protein seq ID NO 1961.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
OS Homo sapiens.
XX
XX WO200157190-A2.
PN
XX 09-AUG-2001.
PD
PE 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSEQ-) HYSEQ INC.
PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;


```

QY 897 KDPGSSGK--ACGEVGGSEPTQTALAEHOKTHPKDGLFTGVCNGLFDRATLKKN 953
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 733 VHLGVHTREKLEKCECGKGFSSQSALEAHQNVHTGEKP-YKCDICDKDFRRHRLTVH 990

RESULT 34
ID ABB59449 standard; Protein; 744 AA.
XX ABB59449;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 5139.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
XX
XX N-PSDB: ABL03552.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX disclosure; SEQ ID NO 5139; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
XX sequences (ABU01840-ABU16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 744 AA:

Query Match 5.9%; Score 311.5; DB 22; Length 744;
Best Local Similarity 18.1%; Pred. No. 3,5e-11;
Matches 193; Conservative 90; Mismatches 281; Indels 505; Gaps 37

QY 90 MDTEHSNPDDSSSGCPDPTWGPERRGEESSGGQLVA-ATGTAAGGGGGLILASPKLG-- 146
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 MPTSSSEISGCGCAIPMLRPSR-----MDQFNMSMAAAAAAAGCGGLGADRNCGS 55
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 147 -----ATPLPESTP-----APPPPPP----- 164
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 56 GGSDBGSGNGNDSSRNSASRISAYETOLAYOQLAGLHGPPPPPSHHREISAFYVVL 115
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 165 ---PPPGVSGHLNPLILEELRVLQOQRIHOMQMTQICHQVLLSGLGQTVGAPASP 221
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 116 PTGKVRPSNSNYEITAMMADKKRELALRE-----AAAAAMLGRRGCGPGCP 163
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

QY	222	SELPCTGAAS	STKRLPL	LESP	SP	IKPA	QTKTTA	SSSSSSSS	SGAE	RRPKA	FHL	YN	PL	GSQ	261
		1	G	-VPP	PGV	-----	-LYGR	-----	-AGVPP	-----	-----	-----	-----	180	
Db	164	G	-VPP	PGV	-----	-LYGR	-----	-AGVPP	-----	-----	-----	-----	-----	180	
QY	262	HFESV	GVGR	SR	HKPT	PAR	SP	AL	CG	STO	OL	AS	PH	LA	FG
		181	-PYLTG	-----	-----	-LGR	SP	TG	AGS	-----	-----	FFPP	CG	AA	AA
Db	181	-PYLTG	-----	-----	-----	-LGR	SP	TG	AGS	-----	-----	FFPP	CG	AA	AA
QY	340	ASPG	LK	-PK	NG	SG	EL	GY	EV	IS	LE	K	PG	R	NC
		220	LD	RLL	RA	CG	RA	S	-----	-----	-----	-----	-----	-----	-----
Db	220	LD	RLL	RA	CG	RA	S	-----	-----	-----	-----	-----	-----	-----	-----
QY	399	KCNV	CG	N	R	F	T	T	CG	N	L	V	H	F	H
		267	SC	D	I	CG	K	A	F	R	Q	H	L	R	D
Db	267	SC	D	I	CG	K	A	F	R	Q	H	L	R	D	-----
QY	459	E	E	E	A	G	I	P	G	G	V	E	K	P	L
		289	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Db	289	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
QY	519	K	A	D	E	N	T	P	G	S	E	G	S	A	I
		289	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Db	289	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
QY	579	L	G	A	S	P	E	T	S	I	K	L	O	L	V
		289	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Db	289	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
QY	639	P	R	A	L	I	N	H	O	H	G	E	R	P	K
		306	S	T	L	V	H	K	T	L	H	E	S	E	S
Db	306	S	T	L	V	H	K	T	L	H	E	S	E	S	-----
QY	699	A	V	T	L	O	H	V	M	H	L	G	O	I	P
		362	N	O	D	L	R	H	A	L	T	H	A	V	E
Db	362	N	O	D	L	R	H	A	L	T	H	A	V	E	-----
QY	727	L	A	O	E	N	S	E	G	O	S	T	A	S	G
		422	S	G	E	K	E	S	E	S	E	M	R	L	K
Db	422	S	G	E	K	E	S	E	S	E	M	R	L	K	-----
QY	775	E	D	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
		477	E	D	E	O	A	E	V	A	L	V	A	F	O
Db	477	E	D	E	O	A	E	V	A	L	V	A	F	O	-----
QY	801	S	G	A	E	E	V	A	T	S	V	A	T	I	V
		537	E	R	G	N	G	I	A	L	P	R	P	S	V
Db	537	E	R	G	N	G	I	A	L	P	R	P	S	V	-----
QY	845	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
		587	P	N	G	D	E	V	L	P	I	L	H	V	R
Db	587	P	N	G	D	E	V	L	P	I	L	H	V	R	-----
QY	888	E	L	N	L	P	-E	A	M	K	K	O	R	E	S
		637	P	L	H	S	P	H	E	A	M	P	S	L	G
Db	637	P	L	H	S	P	H	E	A	M	P	S	L	G	-----
QY	947	R	A	T	E	K	H	M	L						

Query Match	Best Local Similarity	Score	DB	Length	744
Matches	193	Conservative	90	Mismatches	281
Indices	505	Gaps	37		
90	MDTESSNPDPDSSGGPPDPPTWGERGESSGOFVLA-ATGTAAAGCGGGLILASPKTG--	146			
1	MTETSSSSSEISGGGCGALPMLRPSR-----MDQFMMSMAAAAAYGCGGLPGADRNCGS	55			
147	-----ATPLPESTP-----APPPPP-----	164			
56	GSSDGGSONGNGDSRNSASRIASVETLAVOONLAGLHGPPPPPPSHHREISAEPVL	115			
165	---PPPPGVSGHLNPILEELRYLOROHOMQMTQICRQVLILSGIGQTVGAPASP	221			
116	PTGKAVRPGSSNVEIIMAMARKELARE-----AAAAAAMLGRCGGPGGP	163			
222	SELPLGMAASTPRLPLPSPFKPAOTGKTTASSSSSSSSSGAEPKQAFHLYNPLCSQ	281			
164	G-VPPPGV-----LYGP-----AGVPP-----	180			
282	HPSPVSGVGRSHKPTPAPSPALPGSTDOLIASPHLAFPTGTGLLAACL--GAARGLEMA	339			
181	-PLTGT-----PSPSPGAGS-----PPPPGAAAALFLPGLGPGMAG	219			
340	ASBGLLK-PKNGSELGXEYVSISSLEKPGGRHKRCFAKVGSDSALQILHRSHTGERPY	398			
220	LDRRLRAPGRAS-----RPKKQFICKFCNRQFTKSYNLLIHRTHTDERPY	266			
399	KCVVCGNRFTTRGLKLVHFNHREKYPHYQMPNHPVREHLDYVITSSGLPFGMSPPREKA	458			
267	SCDICGAFRQDHLRHD-----RYIH-----	288			
459	EEBAGTGGGVERKPLVASTTALSTESTLLTSTGTSTAVAPGLPTFNKFWLMKAVEPKS	518			
289	-----	288			
519	KADENPPGSEBSAIAGVADSGSATRMQLSKLVTSLPSMALLNLHLSKTSGSFPPLYLEP	578			
289	-----	288			
579	LGASPSSTSKLQQLVERIDROGAVAVASTASAPPTISAPAPSSSASGPNOCVICRLVLSG	638			
289	-----	305			
639	PRALRLHYGNGGERPFCKYCGRAFTSGNLRANHFVGHKTSAPAARAKNSCPICOKKFTN	698			
306	SRTLAIVAKILHMEESPHCRCVCSRSFNRQSNLTKHLTLH---TDHKRYECCSGKVFR	361			
699	AVTLQGVNRMHLGQIIRPGSA-----LSGGG	726			
362	NCDDRRLALTHAVGVNSGDVVDVGEEDENARNLSCQEDSLLLEVDSPPQSPVHNHNGESSG	421			

[illegible]

xx	Sequence	934 AA;	5.8%; Score 309; DB 22; Length 934; Best Local Similarity 19.9%; Pred. No. 6.7e-11; Matches 175; Conservative 80; Mismatches 298; Indels 328; Gaps 33;
2y	161 PPPPPPPVSGSHNLPIPLLEELRVLQORQIHOMQTEQICROVLLLSLQTVGA---	217	
Db	89 PPPPLP-----QVTSHASASAAAAAASNNNAVA	121	
2y	218 -PASSELPGTGASSTKPLPLPSPKPAQTG---KTASSSSSSSSGAE-----	265	
Db	122 VMAASAAAAAASAGGSLP-----PATSGNGGQVTVTTSSSTSGSTSGTT	175	
2y	266 -----PPKQAFPHLHYHLSQHPFVSGVGRSHKPTPAF--SPALGS-----	306	
Db	176 TTAGELMKMEGCIHGVDSGN-----GGNGGQNVALADGPTPIANGTHVCDICGM	229	
2y	307 ---TDQLASPHLAFPTGTGLLAQCLGAARGLAASPGILKPKNGSGELGYEYI---	360	
Db	230 FQFRYQLIV--HRRYHSEKRPFCQVCG--QGFTTSQD---LTRHGKIHIG--GPMETCI	280	
2y	361 -----SLEKPGGRK-----CRFCAKVGSDSALOIHLSHTEBRPKCVNCGN	405	
Db	281 VCFNVFANNSTLERHMKRHSSTDKPFACCTICQTFARKKHLDMHFRSHTEBTPRCQYCAK	340	
2y	406 RFTTGNLKVHFHREKYPH---VOMNPHVPEHL--DYVITSSGLPGMSVPRPEKAE	460	
Db	341 TPTREKHMVNHVAKHGTETHRCDICCKSFTREKELVNHVMHNTGOTPRQOCYCGK---	397	
2y	461 EAGTPGGVERKPLVASTALSTATESLTLLSTGTSTAVAPGLPTENKFLMAVERBSKA	520	
Db	398 -----YTRKEHLAN-----HMSHT	412	
2y	521 DENTPGSGSALAGVADSGATPMQSLKVLTSLSMALLTGHLSKTSFPFYVLEPLG	580	
Db	413 NE-TYPRCE---TCGSEFSKKEH-----FTNH-----LIMHTG	441	
2y	581 ASP-----SET-SKLOQLEKIDROGAVAVASTASGAPTSAPAPSSASGPNQCIVCI	633	
Db	442 EPRHNCDFCSKFTREKHLNHW-RQ-----HNGESPHRSCYCM	479	
2y	634 RVLSCPRLRLHYGONGGERPFCKVCGRAFSTRGNLRAHVEYGHKTSPPARAONSCPTCQ	693	
Db	480 KFTTREKHLVNHTRQHTGETPTFCCTCTCAFTRKD---HMVNHVROHTGESPHKCTYCT	535	
2y	694 KFTNAVTLQOHVYMLGGQIPNGSALSEGGAAQENSSEGSTASGPGSPFQPOQOPS	753	
Db	536 KFTTREKHLTNHVKLTG-----	553	
2y	754 PEEEMSEEEDEEEDVYDEDSLARGSESGEKAISVRGDSSEVSGAEDEVATSVAA	813	
Db	554 -----DSPHKCEYCKQTFRKHEL-----NNHMRQHSNDNPHCVCNCK	592	
2y	814 PTTVAKEMDSNEKARQHTLPPRRPPRDLNHPQMEQGTSDVSGAMEEAKLKGISSPMAA	873	
Db	593 PPTREKHLNHNRSRCHTGR-----	613	
2y	874 LTOEGEGSTPLVEELNLPRAMKKDPGESSGRKACEVCGSPPTOTALREHOKTHPKDPR	933	
Db	614 FTCEFCGKRFPLKGNLLFNRSHTKSGEMERFPACEKCPKNTICKGHLVSHMRSHGGER	673	
2y	934 LETVCFCRQGFIDRAFLKHKMLAHNOV---PPFAPHQ	970	
Db	674 -HACTLCSKAFVERGNLKRHKMKMNHPRDMMPPRPVPH-PQ	712	
RESULT 38			
ABBS9829			
ID	ABBS9829 standard; Protein; 1893 AA.		
KX			
YC	ABBS9829;		

xx	26-MAR-2002 (first entry)	
DT	Drosophila melanogaster polypeptide SEQ ID NO 6279.	
xx		
DE	Drosophila; developmental biology; cell signalling; insecticide;	
xx		
KW	pharmaceutical.	
xx		
OS	Drosophila melanogaster.	
xx		
PN	WO200171042-A2.	
xx		
PD	27-SEP-2001.	
xx		
PF	23-MAR-2001; 2001WO-US09231.	
xx		
PR	23-MAR-2000; 2000US-191637P.	
xx		
PR	11-JUL-2000; 2000US-0614150.	
xx		
PA	(PEKE) PE CORP NY.	
xx		
PI	Venter JC, Adams M, Li PWD, Myers EW;	
xx		
DR	WPI; 2001-656860/75.	
xx		
DR	N-PSDB; ABL03932.	
xx		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more	
PT	genes from Drosophila and for elucidating cell signalling and cell-cell	
PT	interactions -	
PS	Disclosure; SEQ ID NO 6279; 21pp + Sequence Listing; English.	
xx		
CC	The invention relates to an isolated nucleic acid detection reagent	
CC	capable of detecting 1000 or more genes from Drosophila. The invention is	
CC	useful in developmental biology and in elucidating cell signalling and	
CC	cell-cell interactions in higher eukaryotes for the development of	
CC	insecticides, therapeutics and pharmaceutical drugs. The invention	
CC	discloses genomic DNA sequences (ABL01840-ABL30511), expressed DNA	
CC	sequences (ABL01840-ABL16175) and the encoded proteins	
CC	(ABBS7737-ABBS72072).	
CC	The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences.	
xx		
SO	Sequence 1893 AA;	
Query Match	5.8%; Score 309; DB 22; Length 1893;	
Best Local Similarity	18.9%; Pred. No. 1.6e-10;	
Matches	265; Conservative 124; Mismatches 444; Indels 570; Gaps 58;	
2y	6 GSSSLRGPGCEPAERGGDASEHHPOVCARCAQFSDPTFLAH-----QNSC	54
Db	226 GSSPQ--GQCLSSGSEGIGAGDEHMKYLCIPCEVVASAPHEFTNHRCHNYANGDENFT	283
2y	55 C-----TDPVVMYIG-----	65
Db	284 CRICKVYLSSASSLDRHVLVHTGERPNCRCYCHLTFTTNGNMRHMRTHKQHOVASQSQ	343
2y	66 -----QENPS-----	81
Db	344 SQQQQSLQQQQQQSQQRRQQQOHQPSQQQQQNPAAQQQLMGNTLSARAESVDSASCTDVS	403
2y	82 EGHRSQVMDTEHSNPDSGS-----SGPDPWTGPERR-----	115
Db	404 SGSHSRSSSLNNNNNHSKANNLKLDELEVESTEDQDETNKORRLKTTINNIIESE	463
2y	116 -----GESSSGQFLVAATGTAAGGGGGLIASPKTGATPLRPESSTPAPRPPPPPP	167
Db	464 QQEDMDDEADADAVAMLTSTP--DVATILAGASASGA-----ASRSPSPSASPAL	515
2y	168 ---PGVSGS-----HLNPLILEELRVL---QQRQIHOMQTEQICR-Q	204
Db	516 LSCPACGASDFETLPLACVHLDAMHSDIPAKCRDCEVIFATHRQLOSH-----CCRLP	568

QY	205	VLLGSLGOTYGARSP--SELRPGAASTPKLLELPSFKPAQIGKTTAASSSSSSSS	262
Db	569	NALAGGLRPLGLASSPRLNEPEREDENHDEDELL-----EOKERLASQSED----	614
QY	263	GAEPKQAFHLYHPLGSONHPSFVGVSGRSHKPTP-----	APSP 301
Db	615	-----FFHOLYLKHKTANCGAALSHPPSPILKHEADPKDLADIOSILNMTSSSS	663
QY	302	ALPGSTDOLIASPH--LAFPGTTGLLAQSLGAA-----RG-----LEAASPGLK	346
Db	664	SFLRNFEOGSVNPSPNSQOYSLDRDDEEEOADAFTESEFRMKLGEFFPCKLCTAFVPMILA	723
QY	347	PKNGSGELGCVGISLSEKPRGRHKRCFCAKVGSGSALOILRSHNGEPKYNCGNKR	406
Db	724	LK-----GHNVHLGAVGPAQPFRCNMCPYAVCDKALVRHMKTHGDRYEGCAVCNYA	777
QY	407	FTTRGNLKVHF-HRH-----REK-----YHVOOMP	431
Db	778	FTTKANCSERHLLRNHKGKTSREEVKRAIYVYHAPEDACGSDCSRLGEDLADTSPRSISPTP	837
QY	432	HPVP-----EHLDYVITSS-GLPRGMSVPRPKAE	459
Db	838	PPRPVNESKSQLKHMILGENHILAVNQORPLKIQVSTSLDLVKKRSPAPRQOQOQOQOQ	897
QY	460	BEAGT-----PGGVEKRPVASTYALASATSELTLSLTGSTAVAPRGLPTF	505
Db	898	EKSSGALDRSMVDLSDTKRPTGASLTPAVTRP-----TPAAVAP-----	938
QY	506	NKFVLMKAVERKSKADENTPRGSEGSALAGVADSGSATRMO-----LSK	549
Db	939	-----VPRPG-----VTPRLAAAIIEOQOOLLLAQQOLFAGAGEYMOQ	975
QY	550	LVTSLPSMALTLNHLKSTGSPF-FPVVLEPLGASPESTKLOQVLEKIDRQGAVAVASTA	608
Db	976	LFRSL-----MFQSGTGFPRFFPRMAPR---PROANPKRPVWSPPNKRNMPV-GVG	1024
QY	609	SGARTTSABARSSSASGRNQCVTCLRYVLSCPRALRLHYONGBERPKVCYGAFFSTRG	668
Db	1025	VGVVP-----PRGPRKVMVTKNGVL-MPKOKORYRT---ERPRACEHSGARFTLRS	1072
QY	669	NLRAPF-----VGKHTSPAARQAN-----	687
Db	1073	NMERHVKQOHPOFYAORORSNGHHVMRGSGASNVAAAAAATAATVMAAGPSSGFGS	1132
QY	688	-----SCPICQ-----KKTNAVYLOQHVR-----MHL	710
Db	1133	NHHNGHGHSGHSHGNAPRISBOYKCALIQAOLKANKTDLLOALHANGSSSVAAQNPLHF	1192
QY	711	GGQIRP-----GGSALISEGGAA-----QENSSSEOST-ASGPGSPPOQSO	750
Db	1193	GPRLPNPSRPMHNGS--SQNGOQTAMDDDEPKLIDEDENENHVEVAEDVDV-----E	1245
QY	751	QPSRESESEEE-----DEEEEDVTDIDS-----LAGSGSEGE-----	788
Db	1246	EDEDEEMERDEPERLIDDEQRAKEMAEBOELPKLEBQLGTKEAQAQMAETLLEQAIK	1305
QY	789	--KAISVRGDSEEVGAEEVAT-----SVAAPTT-----VKEMDSN	823
Db	1306	ACKRPLSRPTKEVASPAMPRTVAATMOERPAITAASTNSSLTKYIAQAEVYKSLKEVASS	1365
QY	824	E-KARQHTLPPRRPRPDNLDRPMEQST---SDVSGAMEEAKLGTGSSPYMALPDQEG	878
Db	1366	PRKDEQODLVPAAKLYDNAT-SQNMGFNSTFRPSDVANHHQOSE-EDVASGSGSSESNN	1422
QY	879	EGSTPRLVEELNLPEAMKDDPGESSGKACVEVGSQSPPTALBENOKTHPKDPLFTCV	938
Db	1424	SGTEVDYSSSSSEPKKSAVSLAPNVSCPYOQRMFPWSSSLRRHILITHTGQRP-FKCS	1485
QY	939	PCRQGLDRATLKKHMLAHNOV 961	
Db	1483	HCPRLFTTKSCDRHLLRKHN 1505	

Query Match	Best Local Similarity	Matches 143;	Conservative 143;	Score 305.5;	DB 22;	Length 1196;
346 KPKNGSGELTYGEYISLEKPGGRKCRFCACVFGSDAQLQILRSHGTGRPYKACNCGN 405	11	11	11	11	11	11
358 KANLSSPSFYHEHITHGAKR---YECOCGCAFRASQLRVHGNIHTGKPEYCKKCGK 414	11	11	11	11	11	11
406 RFTTGNLKVHFRHREKYP-----HVQKNPNHVPVPEHLIDYITSSGLPYGM 451	11	11	11	11	11	11

[illegible]

XX	DR	XX	WPI: 2001-318749/34.
XX	XX	XX	Primer sets for synthesizing polynucleotides, particularly the 5602
XX	XX	XX	full-length cDNAs defined in the specification, and for the detection
XX	XX	XX	and/or diagnosis of the abnormality of the proteins encoded by the
XX	XX	XX	full-length cDNAs -
XX	XX	XX	Claim 8: SEQ ID 17104; 2537pp + CD ROM; English.
XX	XX	XX	The present invention describes primer sets for synthesizing 5602
XX	XX	XX	full-length cDNAs defined in the specification. Where a primer set
XX	XX	XX	comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX	XX	XX	to the complementary strand of a polynucleotide which comprises one of
XX	XX	XX	the 5602 nucleotide sequences defined in the specification, where the
XX	XX	XX	oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX	XX	XX	of an oligonucleotide comprising a sequence complementary to the
XX	XX	XX	complementary strand of a polynucleotide which comprises a 5'-end
XX	XX	XX	sequence and an oligonucleotide comprising a sequence complementary to a
XX	XX	XX	polynucleotide which comprises a 3'-end sequence, where the
XX	XX	XX	oligonucleotide comprises at least 15 nucleotides and the combination of
XX	XX	XX	the 5'-end sequence/3'-end sequence is selected from those defined in
XX	XX	XX	the specification. The primer sets can be used in antisense therapy and
XX	XX	XX	in gene therapy. The primers are useful for synthesizing polynucleotides,
XX	XX	XX	particularly full-length cDNAs. The primers are also useful for the
XX	XX	XX	detection and/or diagnosis of the abnormality of the proteins encoded by
XX	XX	XX	the full-length cDNAs. The primers allow obtaining of the full-length
XX	XX	XX	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX	XX	XX	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX	XX	XX	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX	XX	XX	represent oligonucleotides, all of which are used in the exemplification
XX	XX	XX	of the present invention.
XX	XX	XX	Sequence 725 AA:
XX	XX	XX	Query Match 5.7%: Score 304; DB 22; Length 725;
XX	XX	XX	Best Local Similarity 20.5%: Pred. No. 1e-10;
XX	XX	XX	Matches 181; Conservative 78; Mismatches 275; Indels 348; Gaps 37
QY	142	SPKIGATPLPE--STPAAPPPPPPPPPGVSGHINIPLELRYLQOROHOMQTE 199	
DB	107	SPK--KELPAEESPVLGPDSDEPRRPQ-----RAAQ----- 135	
QY	200	QICROYALLGSLGQTVGAPASPSF--LPETGAASSTKPLPLPSPIKPAOTGKTTASSS 257	
DB	136	-----PAEPGQWRLP-----PSSKQPLSP-----GPKRTFQALQESSP 168	
QY	258	SSSSSAAE-----PRKQAFPLLYPLG-----SQHPFSGVGGR 291	
DB	169	QGPSPWDEESSRDOELAVLCLTFEDYVENKAW--PAHPLGSGSRTPDKKEFKQOEBPKG 226	
QY	292	SHKPTP-----APSPALPGS--TDQLTASPLPLAFPTGTLIAAOCLGAAGLMAAASPG 343	
DB	227	AAMPPTPLAASQADSPCVGPBPSCAOSLGGGAASGGEDESL--LGSSEILEYKVAEG 282	
QY	344	LKPKNGSGELGY---GEVVISLSEK-----PGGR-HKRCFCAYFGSDAIDH 388	
DB	283	VPEP---NPELOFCADCGVGFPPQLSRLLKAHOLRSHPAGRSFCLCGKSGRSISILKH 339	
QY	389	LRSHTEGPRPKCNCVGNRFTRGNLKVFNHRRREKRPYQOMNHPRVENDLYVTTSSLP 448	
DB	340	MRNHTDERPACHLCGHRFRSSSHLSKILLTASSF-----PAFL----- 378	
QY	449	YGVSVPEPKAEDEAGTGGCGVERKPLVASTALSATESILLSTGTSTAVAPGLTFENKF 508	
DB	379	-----CAEGKRGFQRRASLYOHLLAHADQ----- 403	
QY	509	VLMKAVERKSKADENTPGSGESALAGVADSGSATFMOLSKLTVSLPMAILLTNHLKSTG 568	
DB	404	--KRPCAPESKAE--APPPLTD-----VLGSH----- 425	
QY	569	SPFPFVYVLEPLGASPSSETSKLQDLVEKIDRGCAVAVASTASGAPPTTSAPADSSASGPNQ 628	

Db 426 -----CGSFFORRSSLKRLH-RIHARDKDRRSSESGS-----RRDSDRRPFV 468
QY 629 CVICLVLSCPALRLHLHYGOHGEPRPFCKXVCGRAFPSTNGNLRAHFGHKTSPAARQNS 688
Db 469 CSDCGKAFRSEHLVAHRRVHTGERPFSCQACGRSFTQSSQL---VSHQRVHTGEKPYA 524
QY 689 CPICOKFTNAVTLQOHVRMHLGQIPNGSALSSEGGAAQENSSEOSTASGPGSPQPQ 748
Db 525 CPQCGKRFRVRRASLARHLTH-GGPRPHCTQCGKSFQGTQDLARHQRSHTG----- 575
QY 749 SQQSPSEEESEEEDEEEDVTDEDSLAGRSESGGEKAI SVRGDSEEVSGAEEVA 808
Db 576 -EKPC-----RCSECG-----EGFSQSAHLARHQ-- 598
QY 809 TSVAAPTIVKEMDSNEKAPQHTLPRPPRPDLHPQPMEGTSDVSGAMEEAKLEGIS 868
Db 599 -----RIHTGEK-----PHACDTGCH-----RFRNSS 620
QY 869 SPMAALTOEGEGSTPLVEELNLPKAMKKDPGSSGKRACEYCGGSPPTOTALSEHOKTH 928
Db 621 -----NLARHRSHTGERP--YSCQTCGRSFRRAHLRHLATH 657
QY 929 PKDG-----PLFTCVFCRGFLDRATLKHHMLAHNOVPPFA 965
Db 658 AEPGOEQABEPPOECVECGKSFSRSCNLRH-LIVHTGARPY 698

Search completed: January 13, 2003, 15:20:09
Job time : 54.9432 secs

TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-399-411-2

Query Match 4.7%; Score 249.5; DB 2; Length 1706;
Best Local Similarity 22.1%; Pred. No. 2e-08;
Matches 186; Conservative 69; Mismatches 281; Indels 307; Gaps 38;

14 PCGEPARGGDASEEHHPQVCAK-----CCAQ-----FSDPTFLAHQNSCTDP 58
851 PCD---SEGKEFEKNHIAQPAKAKKPTTCMLQKVLNENGVSLPTE-----TTP 898
59 PVMYIIGQENPSN-SSASSAPRPE-GHSRSQVMDTEHSNPPDSGSGPPDPPTWGPERRG 116
899 EV-----TRSPSPCKSPDTPQDPDELGPDSGCVPTAESPPVEVGGSPPLQT-----A 946
117 EESSGQFLVATGTAAGGGGGLILASPKLGATPLPESTPAPPPTPP-----PPP 167
947 SLSSGQ-----LPLLTPTTBSPPPCPPVLTVAATPPP 979
168 PGVSGGHLNPLILEELRYLQORQHOMQMTQICRQVLLGSLGQTVGAPASPELPGT 227
980 PLPLTVPLSHP-----SSDASPQCCPSP 1002
228 GAASSTKPLPLPS-----PIKP---AOTGKTTASSSSSSSSGAEPPQAF 272
1003 FSNHTAOSPPLISTVSPSPSPPIPVPEPLMSAASPOPPILSSSSSSSS-----1052
273 HLHYPLSGQHPFVSGVGRSHKPTPAPSPALP---GSTDLILAS-PhLAFPGTGLLAAQ 328
1053 ---FPSSS-----CSSTSPSPPLSAVSVSSGMDLEASLPAYTFK-----1091
329 CLGARGLEAASPLGLKPKNGSGELGYEVISLSEKGRHKRCFAKAVGSDALQIH 388
1092 -----OEBESSEG-LKKEKERRPPAGGSVVG-ETFSKNFTCNCESPPLSKDKLKH 1141
389 LRSHTGERPYKCNVCNRPRTTGNLKVH-FHRR-----REKYPHV-QMNPFPVP 435
1142 LSVNAEWPKEFCFCVQLKEVKTDLSEHRRFLHGVGNIFFCVSKCKKEPAFLCNLQHQRD 1201
436 EHLDYVITSSGLPYGMSVP-----PKAEDEAGTPGGGVRRKPLVASTTALSAIESLTLL 490
1202 LHPDEVCTHNEFESGTLRQNTDPSKAVEN-----1233
491 STGTSTAVAPGLPTNKKFVLMKAVER-KSKADENTPPGSEGSATAGVADSGSATRMQLSK 549
1234 -----MPLSLPE-----EGLETSREELNDSSSE-----ELVTTIKIMASG 1267
550 LVTSLPSNAL-LTNILKSTGSPFPR-VLEPLCASPSETS-----KLDQLV 593
1268 IKTDQDVRVLGNQHYPSKRRPFOYHNRRPMGIGVATATFTTNIPOTFTTATIRCTCG 1327
594 EKIDRQ-----AVAVSTASGAPTTSAFARSSASAGNOVICI-----RVLSC 638
1328 KGVDMPELKHILILACASASDKKRYPRKKNPVRLKQTVQKNGVVVLNDSGKNAPFRRGG 1387
639 PRALRLHGGHNGGERPFCKVC-----GRAFSTRGMLRAHVGHTKTSRA 682
1388 PKRLSFNV-ELGKMSPNKLTALAKKKKNOLVQKALLOKNRAAKAKALR-----DTSEA 1440
683 ARAONSCPIQCKFTNAVILQOHVRRMLGQIPNG-----SALSEGGAOENSESOSTA 738
1441 ---SSHTCYCQDEFTYIGSLNKHAAFSCKRPLSPKRYKSHSKKGHASSSSSDRRNS 1498
QY 739 GGP 741
Db 1499 GHP 1501

RESULT 6
US-08-516-859A-2
Sequence 2, Application US/08516859A
Patent No. 6069231
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/516,859A
FILING DATE: 18-AUG-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1776
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-516-859A-2
Query Match 4.7%; Score 249.5; DB 3; Length 1706;
Best Local Similarity 22.1%; Pred. No. 2e-08;
Matches 186; Conservative 69; Mismatches 281; Indels 307; Gaps 38;
14 PCGEPARGGDASEEHHPQVCAK-----CCAQ-----FSDPTFLAHQNSCTDP 58
851 PCD---SEGKEFEKNHIAQPAKAKKPTTCMLQKVLNENGVSLPTE-----TTP 898
59 PVMYIIGQENPSN-SSASSAPRPE-GHSRSQVMDTEHSNPPDSGSGPPDPPTWGPERRG 116
899 EV-----TRSPSPCKSPDTPQDPDELGPDSGCVPTAESPPVEVGGSPPLQT-----A 946
117 EESSGQFLVATGTAAGGGGGLILASPKLGATPLPESTPAPPPTPP-----PPP 167
947 SLSSGQ-----LPLLTPTTBSPPPCPPVLTVAATPPP 979
168 PGVSGGHLNPLILEELRYLQORQHOMQMTQICRQVLLGSLGQTVGAPASPELPGT 227
980 PLPLTVPLSHP-----SSDASPQCCPSP 1002
228 GAASSTKPLPLPS-----PIKP---AOTGKTTASSSSSSSSGAEPPQAF 272
1003 FSNHTAOSPPLISTVSPSPSPPIPVPEPLMSAASPPPTLSSSSSSSS-----1052
273 HLHYPLSGQHPFVSGVGRSHKPTPAPSPALP---GSTDLILAS-PhLAFPGTGLLAAQ 328

1053 ---FPSS-----CSSSTSPSPPLSAVSSVSSGDNLEASLPAVTEK----- 1091
329 CLGAARGLAASPGILKPKNGSGELGYEVISLSLEKPGGRHKCRFCAKVGSDSALOIH 388
1092 -----QEBSESEG-LKPKKEAPRAGGOSVQ--ETFSKNFICNVCESPLSTIKDLTKH 1141
389 LRSHTGERPKYKCNVCGNRFTRGNLKVH-FHRH-----REKYPHY-QMNPHEVP 435
1142 LSVHAEMWPKCEFCVOLEFKYKTDLSEHRFLHGVGNIFVCSVCKKEFAFLCNLOOHORD 1201
436 EHLIDYITSSGLPYGMSP-----PEKAEAGTPEGGVBRKLVASTALASATESLTL 490
1202 LHPDEVCTHHEFESGTLRPONTDPSKANVEH----- 1233
491 STGTSTAVAPGLPTFNKFLVLMKAVEP-KSKADENTPPGSEGSALAGVADSGSATRMOLSK 549
1234 -----MSLPE-----EPLETSREELNDSS-----ELYTTIKIMASG 1267
550 LVTSLSFMSAL-LTNHLKSTGSPRPY-VLEPLGASPSETS-----KLQOLV 593
1268 IKTKDDPVRILGNHYPSEFKRPFOYHHRNPMGIGVATNFTTNINIPQFTTALICTKCG 1327
594 EKIDRG-----AVAVASTASGAPTTASAPSSSASGPNOCVYCL-----RVLSC 638
1328 KGVDMPELKHILLACASASDKKRYTPKKNPVPKQTVQPKNGVVLNDSGKNARRMOQ 1387
639 PRALRLHYGOHGERPFCKVC-----GRAFTRGNLRAHFVGKHTSPA 682
1388 PKRLSFNV-ELGKMPNKLKLSLAKKKNOVLQKAILKNNRAKQKADR-----DTSBA 1440
683 ARKONSPICOKKTTNAVTLDOHVRNHLGGQIPNG-----SALSEGGAOENSSEOSTA 738
1441 --SSHICPYCDREFTYIGLSLNKHAAPSCPKPLSPSKRVKSHSSKGGHASSSDSSDNSS 1498
739 SGP 741
1499 CHP 1501

RESULT 7
US-09-586-472-2
Sequence 2, Application US/09586472
Patent No. 632335

GENERAL INFORMATION:

APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting Zinc Finger Proteins

NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/586,472
FILING DATE: 01-Jun-2000

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/528,706
FILING DATE: 17-MAR-2000
APPLICATION NUMBER: US 08/516,859
FILING DATE: 18-AUG-1995
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
APPLICATION NUMBER: US 08/292,683

FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 4130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-586-472-2

Query Match 4.7%; Score 249.5; DB 4; Length 1706;
Best Local Similarity 22.1%; Pred. No. 2e-08;
Matches 186; Conservative 69; Mismatches 281; Indels 307; Gaps 38;

14 PCGEPRERGDASEEHNPQYCAK-----CCNQ-----FSDPTFLAHQNSCCTDP 58
851 PCD---SEGKEFENHLAOPAKKKKPTTCMLQKVLNEYNGVSLPE-----TTP 898
59 PYAVITIGQENPSN-SSASAPRPE-GHSRSQVMDTEHSNPPDSGSGPDPPTWGPERRG 116
899 EV-----TRSPCKCKPDPDPDELGPDSGCVPTAESPEVVGPPSPLOT-----A 946
117 ESSSGFLVAATGTAAGGGGLILASPKLIGATPLPESTPAAPPPPPP-----PP 167
947 SLSSGQ-----LPLLTPTPESSPPCPDVLTVANPP 979
168 PGVSGHLNIPILIELKVLQROIHOMQMTBQICQVLLIGLGYGVAAPASELPGT 227
980 PLLPTVPLSHP-----SSDASPPQCCSP 1002
228 GAASSTKPLPLPFS-----PIKP-----AQTGKTASSSSSSSGAEPKQAF 272
1003 FSNTTAQSPLILSPVSPSPPIPPVEPLMSAASGPPILSSSSSSSS----- 1052
273 HLXNPLGSOHPFSVGGVGRSHKPTAPSPALP-----GSTQDLIAS-PHLAPFGTTGLAAQ 328
1053 ---FPSS-----CSSSTSPSPPLSAVSSVSSGDNLEASLPAVTEK----- 1091
329 CLGAARGLAASPGILKPKNGSGELGYEVISLSLEKPGGRHKCRFCAKVGSDSALOIH 388
1092 -----QEBSESEG-LKPKKEAPRAGGOSVQ--ETFSKNFICNVCESPLSTIKDLTKH 1141
389 LRSHTGERPKYKCNVCGNRFTRGNLKVH-FHRH-----REKYPHY-QMNPHEVP 435
1142 LSVHAEMWPKCEFCVOLEFKYKTDLSEHRFLHGVGNIFVCSVCKKEFAFLCNLOOHORD 1201
436 EHLIDYITSSGLPYGMSP-----PEKAEAGTPEGGVBRKLVASTALASATESLTL 490
1202 LHPDEVCTHHEFESGTLRPONTDPSKANVEH----- 1233
491 STGTSTAVAPGLPTFNKFLVLMKAVEP-KSKADENTPPGSEGSALAGVADSGSATRMOLSK 549
1234 -----MSLPE-----EPLETSREELNDSS-----ELYTTIKIMASG 1267
550 LVTSLSFMSAL-LTNHLKSTGSPRPY-VLEPLGASPSETS-----KLQOLV 593
1268 IKTKDDPVRILGNHYPSEFKRPFOYHHRNPMGIGVATNFTTNINIPQFTTALICTKCG 1327
594 EKIDRG-----AVAVASTASGAPTTASAPSSSASGPNOCVYCL-----RVLSC 638
1328 KGVDMPELKHILLACASASDKKRYTPKKNPVPKQTVQPKNGVVLNDSGKNARRMOQ 1387
639 PRALRLHYGOHGERPFCKVC-----GRAFTRGNLRAHFVGKHTSPA 682
1388 PKRLSFNV-ELGKMPNKLKLSLAKKKNOVLQKAILKNNRAKQKADR-----DTSBA 1440

OY 683 ARAQNSCPIQCKFTNAVTLOQHVHMLGGQIPNG-----SALSEGGAQENSSEOSTA 738
Db 1441 --SSHICPCDREFTYIGSLNKHAAFCSPKRPKPLSPSKRKRVSHSSKKGGHSSSSSDNNSS 1498
OY 739 SGP 741
Db 1499 CHP 1501

RESULT 8

US-09-528-706-2
; Sequence 2, Application US/09528706
; Patent No. 6468985
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/528,706
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/516,859
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1776
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ. ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1706 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-528-706-2

Query Match 4.7%; Score 249.5; DB 4; Length 1706;
Best Local Similarity 22.1%; Pred. No. 2e-08;
Matches 186; Conservative 69; Mismatches 281; Indels 307; Gaps 38;

OY 14 PCGEFAERBGDAEENHPQCAK-----CCAQ-----FSDPTFLAHONSCCTDP 58
Db 851 PCD---SEKEFEKNHIAQAAKKKKPTTCLMLQVLLNEYNGVSLPTE-----TTP 898
OY 59 PVMVILIGGOENFSN--SSASSAPPE--GHSRSQVMDTEHSNPPDSSSGPPDPTWGPERRG 116
Db 899 EV-----TRSPSPCKSPDQRPDELPDSSCSVPTASPEPVVGPSSPLQT-----A 946
OY 117 EESSGCLVAATGTAAGGGGLILASPKLGATPLPEESTPAPPPPPP-----PPP 167
Db 947 SLSSG-----LPPLLTPTSPSSPPPCPPVLTVAATPPP 979
OY 168 PCVGSGLNIPILIELRLVLOQRIOHOMQMTQICROVLLGLSGLOTVAPASPSLPGT 227

Db 980 PLLPTVPLSHP-----SSDASPQCCSP 1002
OY 228 GAASSTPLPLPFS-----PIKP-----AQTGKTTASSSSSSSSGAEPPKQAF 272
Db 1003 FSNMTAOSPPLPLSPVSPSPSPPIPPVEPLMSAASGPPTLSSSSSSSSSS----- 1052
OY 273 HLVPILGQHPFSVGVGRSHKPTAPSPALP---GSTDOLIAS--PLLAPPGTGLLAQ 328
Db 1053 ---FPSS-----CSSTSPSPPLSAVSSVSSGDLNLSLPAVTK----- 1091
OY 329 CLGAARGLEAASBPGLKPKNGSGELGYGEVISTLEKPGGRHKRCFCAYGSDSALQIH 388
Db 1092 -----QEESESEG-LKPKEAPAGGQSVQ--ETFSKNFLCNCCESPFLSKDLTKH 1141
OY 389 LRSHTGERPYKCNVCGNFTTRGNLKVH-FHRH-----REKYPHV-QMNPHPVP 435
Db 1142 LSVHAEMPFCEFCVOLFVKYKTDLSEHFLHGVGNIFVCSYCKKEFAFLCNQOHRD 1201
OY 436 EHLDYVITSSGLPYGMSVP-----PEKAEAGTPEGVERKFLVASTLMSATESLTL 490
Db 1202 LHPDEVCTHFEESGTLRPQNFDPKNAVEH----- 1233
OY 491 STGSTAVAPGLPTFNKFLMKAVEP-KSKADENTPPGSEGSIAGYADSGSATRMQLSK 549
Db 1234 -----MPSLPE-----EPLTSREELNDSE-----ELYTTIKIMASG 1267
OY 550 LVTSLSPMAL-LTNHLSTGSEPPPY-VLEPLGASSETS-----KLOQLV 593
Db 1268 IKTKDPVRLGLNQHYPSEKPPFPQYHHRNPMGIGVATNFTTHNIPQFTTARCTKCG 1327
OY 594 EKIDROG-----AVVASTASGAPTSAPAPSSASAPNQCIVL-----RYLSC 638
Db 1328 KGYDNMELKHILLACASASADSKKRTPKKNPVPLKOTVOPKNGVVDLNSGKNAFRMGQ 1387
OY 639 PRALRLHYGGHGERPPKCYC-----GRAESTRGNLRAHFVGHKTSPPA 682
Db 1388 PKRLSPNV-ELGKMSPNKLKLSALKKKNOVLQKAILQKRAAKQKDLR-----RYLSEA 1440
OY 683 ARAQNSCPIQCKFTNAVTLOQHVHMLGGQIPNG-----SALSEGGAQENSSEOSTA 738
Db 1441 --SSHICPCDREFTYIGSLNKHAAFCSPKRPKPLSPSKRKRVSHSSKKGGHSSSSSDNNSS 1498
OY 739 SGP 741
Db 1499 CHP 1501

RESULT 9
; Sequence 4, Application US/08459568
; Patent No. 5811304
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,568
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

US-08-459-568-4
; Sequence 4, Application US/08459568
; Patent No. 5811304
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,568
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:


```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-399-411-4

Query Match      4.5%; Score 236; DB 2; Length 1719;
Best Local Similarity 18.1%; Pred. No. 1.6e-07;
Matches 230; Conservative 137; Mismatches 407; Indels 494; Gaps 50;

OY 27 EEHHPOVCAKCAQFSDPTE--FLAHQNSCC-----TDDPV----- 60
D 77 EYYVYPL-GMCMCIDATPEKGNMLRYVMACSGEONLFPLEINRAIYYTTLPIAGEE 135
OY 61 -WYIGQENP-----SNSASAPRPGHSRSQ-----VMDTE-HSNP 98
D 136 LLYWYNGEDNPETLAAIEEERASARSKRRSPKSRGKKKKOEKNKGNKIQDIQLKTS 195
OY 99 DSGSSPPDDPTWGERGEESSGOFLVAATGTAAGGGGLIASPILGATPLPE-STPA 157
D 196 DFTSAMNRDSEGGKEDEKPSASALEQPA-----TLQEVASQGEVPELATPA 243
OY 158 PPPPPPPPPGVSGHLNLPILILELRVLQORQHOMQTEQICQVLLGSLQTVGA 217
D 244 PAMEPQPEPERLEAACEVNDLGE-----EEEEDEDEEE----- 282
OY 218 PASPELPGTGAASSTPKPLPLFSPRIKPAQTGKTAASSSSSSSGAEPKQAFHLYHP 277
D 283 DDDDELEDEGEERASMP-----NENSVKEP----- 309
OY 278 LGSQHFFSVGVGSHKPTRAPSPALPSTDLIASPHLAFPTGTGLLAQCLGAAGLE 337
D 310 -----INCKEP-----EDLLEPKTSEET-----LE 332
OY 338 --AASPGLPKNGSGELGXEYISLEKPGRHKRCFAKVFSGDSALOIHLSH--T 393
D 333 DCEVYPRAMQIPR--TKEANGDVETETMF-----CQCKERTTKQGLERHMHITST 385
OY 394 GERPYICNVCGNRTTNGNLVHFRHR--EKYPVQMNPHVPEHLDVYITSSGLPYG 450
D 386 VNHAFKCYCKGKAFGTQINRRHRHRHAGLKRKPSQTLQP--SEDL-----ADGRASG 437
OY 451 MSVPKEAKEEAGTGGGVEKRLVASTTALSATESLTLLSTGTSTAVAP-----GL 502
D 438 ENV---ASKDDSSPSPSGPDLINMSERASODTINSSVVEENGVEKELHCKYCKVGT 494
OY 503 PT-----FNKFLVLMKAVERPSKADENTPPGSEGSALAGY----- 536
D 495 HTMMRRHQRRVHEHLLRPKGVRRKGLEEROPRAEOAQATQNVYVSTEEDEEADVY 554
OY 537 -----ADSGATRMQLSKLVTSLPMSALLTNH 563
D 555 IMDISNISENLNYIDGKIQTNNNTSNCDVIEMESASADLVGINCLLT--PYTVEITQN 612
OY 564 LKSTGSFPFPYVL--EPLGASPS----- 585
D 613 IKTT-QVPTEDLPKEPLDGTSTNSEKKRRRTASPPALPKIKAETSDPMVPSCSLSPLSI 671
OY 586 -----TSKIQLOLVE-----KIDROGAVAVASTAGAPTS 615
D 672 STTEAVSFHKEKSVYSSKLKQLLOTOKLTPPAGISATEIAKLGVCVASAPASMLVTS 731
OY 616 APAPSSASASPNOVCICLRYLISCPRALRLHYGOHG-GERPFKCVKGRAESTRGNL--R 671
D 732 SREKRRRTSSP-----SSP-----QHSFALRDFGKPSDGKAAWTDAGLSKK 773
OY 672 AHFVGHKTSFA-----ARAQNSCPIC-----OKKFT-----NATYLOOHVNRHLGGQI 714
D 774 SKLESHSDSPAMSLSGDERETVSPPCDEYKMSKEWTASASAVSCNOQPL-----DL 827
OY 715 PNGSALSEGGAA-----OENSSQSTASC----- 740
D 828 SSGVKKKAEGTGKTPYQWESVLDLSVHKHCHSDSECKEFKESHSHSVQPTCSAVKKRKPPTC 887
OY 741 -----PGSFPOPOSOQSPSEEMSEBEED-----EEEEEDVTD 774
```

```

D 888 MLQKVLINXNGIDLPVENPAPADGTRSPSPCKSLAQPDPLDGGSGFPAPVTESTPDVCP 947
OY 775 EDLSIAGSGSESGE---KAISVRGDSSEVSGAEAEVAT-----SVAAPT 816
D 948 SSPALQTPSSLSGGLPPLLIPTDPSSPPCCPPVLTVATPPPLLPVPLPAPSSASBP 1007
OY 817 VKEMDSNEKAPQHTLP-----PPPPP----- 838
D 1008 CPSPLSMATA-QSPPLPILSPVSPSPSPPIPVPEPLMSASAPGPTLSSSSSSSSSFS 1066
OY 839 ----DNLDPQMEQTSVDVSGAMEEAKLEGISSPMALTOEG---EGTSPLYEELN 890
D 1067 SSSSSSPSPPLSAISSVSSGDNLEASLPMISFKOELEENGILKPREEQSAAEQDVV 1126
OY 891 LPEAKMDPGSSSGKKAQEVCGQSFPTOTALBEHOKTHPRDGLFTVCVFCROGLDRATL 950
D 1127 VQETFNKN-----FYGVNCSPPLSITKDLTKHLSIHHEWP-FKCFVQLFKDKTDL 1178
OY 951 KKHMLAH 958
D 1179 SEHFFLH 1186
```

```

RESULT 11
US-08-516-859A-4
; Sequence 4, Application US/08516859A
; Patent No. 6069231
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/516,859A
; FILING DATE: 18-AUG-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1776
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9901
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-516-859A-4

Query Match      4.5%; Score 236; DB 3; Length 1719;
Best Local Similarity 18.1%; Pred. No. 1.6e-07;
Matches 230; Conservative 137; Mismatches 407; Indels 494; Gaps 50;
```

27 EEHHPQYCAKCCAFSDPTE--FLAHQNSCC-----TDPV-----60
177 EYVYPNL-GWMCIDATDEPKGMWLRVYNNACSGEONLPLEINAIYKTLKPIANGEE 135
61 -WYIGGOENP-----SNSASAPRPEHRSQ-----VMDTE-HSNPP 98
136 LLYWYNEDNEIIMAAIEERASARSKRSSPKSKGKKKSOENKKNKIDQIQLTSEP 195
99 DSSSSGPPDPWGPERRGESSGQPLVAATGTAAGGGGLLASKLCLATPLPE-STPA 157
196 DFTSANMRDASGPEDEKPPASALEQPA-----TLQEVASQEVPPLEATPA 243
158 PPPPPPPPPPPGSGHINIPLEELRVLQORQIHOMQTEQICQVLLLSIGTVGA 217
244 PAMRQOPRPERDLRLAACEVNDLGE-----EEEEDEDEEE-----282
218 PASPELPGTGAASSTKPLPLPSPIKPAQTGTTASSSSSSSSGAPPKQAFHLYHP 277
283 DDDDELEDEGEESAMP-----NENSVKRPE-----309
278 LGSHPSPVSGVGRSHKPRPASPALPGSTDOLASPHLAPPTTGLLAOCLGARGLE 337
310 -----IRDEKP-----EDLLEPKTTEET-----LE 332
338 --AASPGLKPKNGSGELGYGEVVISLEKPGGRHKCRFCAGVGSALQHLRSH--T 393
333 DCSGVYTPAMQIPR--TKREANDVFETPMR-----COHCRKFTTKOGLRHMHHIST 385
394 GERPKCNVCNRRFTTGNLKVHFHRH--EKYPHVOMNHPVPEHLDYVITSSGLPYG 450
386 VNHAFKCKYCKCAFGTQINRRHRHRHEAGLKRKPSQTLQ-----SEDL-----ADGKASG 437
451 MSVPPEKAEERAGTPGGVERKPRIVASTALSTATESLTLSTGTSTAVAP-----GL 502
438 ENV--ASKDDSPSLPBDCLIMNSEKASQDTINSSVVEENGVELHPCKYCKKVFGT 494
503 PT-----FNKFLVLMKAVPEPKSKADENTPPGSESGAIGV-----536
495 HTNMRHRQRARHERLIRKGVARKGLLEQPPAEOQAOTQNYVPSTEPEEGEADVY 554
537 -----ADSGATRMQSLKLVTSLSWALLTNH 563
555 IMDISSINISEMILNYIDKIQTNNTNNDVIEIESASADLYGINCLLT--PVTEITON 612
564 LKSTGSPFPVVL--EPIGASSE-----585
613 IKTT-QVPTVEDLPKEPLGSTNSEAKKRRTASPPALPKIKAEITSDPMVPPSCSLPLSI 671
586 -----TSKLOQLE-----KIDROGAVAVASTAGAPTTS 615
672 STEEAVSFHKRKSIVLSKLLQLODKITPPAGISATETAKLQPCVSAASMLPVTS 731
616 AAPRSSASGPNQVCYICRLVLSCPRALRLHYGHG--GERPKCYCGRAFTSTRGNL--R 671
732 SREKRRRTSSPP-----SSP-----OHSPALRDFGKPSDKRAAMWDAGLTSKK 773
672 AHVVGHTSPA-----ARAONSCRIPT-----QKFT-----NAVITQOAHMHLGGQI 714
774 SKIESHSDSPAMWSLGRDERETVSPCPDEYKMSKWTASSAFSSVCWOQPL-----DL 827
715 PNGSALSEGGA-----OENSSEOSTASG-----740
828 SSGVKQAKAEGKTPVQMESVLDLSYHKKHCSSEGEKFKESHVQPTCSAVKRRKPTTC 887
741 -----PGSEPOQSOQPSPEEMSEEEED-----EEEEVDYD 774
888 MLQKVLINXYNGIDLPVENPADGTSPSPCKSLAQPDPDLGPGSGFPAPYESTPVDOP 947
775 EDDLAGSGSEGG-----KAISVGRDSEEVSGAEEVAT-----SVAAPT 816
948 SSPALQTPSSISGQLPPLLIPTDPSSPPCPVLYATVATPPPLLPYVLPAPSSSASPH 1007

QY 817 VKEMDSNEKAPQHTLP-----PPPPP-----838
Db 1008 CQSPLSNATA-QSPLPLISPVSPSPSPIPVPEPLMSAASPGPPTLSSSSSSSSSSPS 1066
QY 839 ----DNIDHPQMEQGTSDVSGAMEEAKLEGISSPMALITQEG----EGISTPLVEELN 890
Db 1067 SSSSSSPSPPLSAISSVSSGDNLEASLPMISFQOELENEGLKPREEQSAEAEOVY 1126
QY 891 LPEAMKDPGSSGRKACVCGQSPPTOTALPEHQKTHPKDGPFTGVFCRQGFDRATL 950
Db 1127 VQETPNK-----FVCNVCBSPFLSTKDLTKHLSTHAERWP-FKCFVCQVLFKDKIDL 1178
QY 951 KKHMLAH 958
Db 1179 SEHRFLH 1186

RESULT 12
US-09-586-472-4
; Sequence 4, Application us/09586472
; Patent No. 632335
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; Zinc Finger Proteins
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/586,472
; FILING DATE: 01-Jun-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/528,706
; FILING DATE: 17-MAR-2000
; APPLICATION NUMBER: US 08/516,859
; FILING DATE: 18-AUG-1995
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 4130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-586-472-4

Query Match 4.5%; Score 236; DB 4; Length 1719;
Best Local Similarity 18.1%; Pred. No. 1.6e-07;
Matches 230; Conservative 137; Mismatches 407; Indels 494; Gaps 50;
QY 27 EEHHPQYCAKCCAFSDPTE--FLAHQNSCC-----TDPV-----60

```

77 EYVYRNLL-GWMCIDATDPEKGNMLRYVWMAACSGEONLPLEINRAIYYKTLKPIAPGEE 135
1Y -MVIIGGQENP-----SNSASSAPRPGHRSQ-----VMDTE-HSNPP 98
136 LLLVWYNGEDNPETIAAIEERRASARSSPKSRKGGKKSQENKNGKNIQDIQLKTSERP 195
99 DGGSSGPPDPTWGPERRGESSGOFVLAATGTAAGGGGLILASPLKATPLPPE-STPA 157
196 DFTSAMKRDASBEGKEDEKRSASALEOPA-----TLOEASGEVPELATAPA 243
158 PPPPPPPPPVGGSGHNLPLLEELRVLDQROJHOMQTEQICROVLLGSLGTVGA 217
244 PAMEPQPEPDERLEAACEVNDLGE-----EEEEEEDEEE----- 282
218 PASPSPLCTGAASSTPLPLPLFSPKPAQTKTASSSSSSSSSSGAEPKQAFHLIYHP 277
283 DDDDELEDEGEERASMP-----NENSVKEPE----- 309
278 LGSQHPPFVGVSGRSHKPTPARSPALPGSTQDILASPHLAFPTGTGLLAQSLGARGLE 337
310 -----IRCDER-----EDLLEPKTTSEET-----LE 332
338 --AASPGLKPKNGSELGYGEVYISLEKPGRHKRCFAKVFSGDSALQIHLRSH--T 393
333 DCEVYPRAMQIPR--TKEANGDVETETMF-----CQHCERKFTTQGLEBHHIHIST 385
394 GERPYKCNVCNRRTTKGNLKVHFRHR--EKYHYQMNPRHVPRELDYVITSSGLPYG 450
386 VNHAFFKCKYCGKAGTQINRRRHRHREAGLKRKPSQTLP--SEDL-----ADGRASG 437
451 MSVPRKEAEEBAGPGGVERKRPVASTALSTATESLTJLSTGTSTAVAR-----GL 502
438 ENV---ASKDDSPPLSGPDLIMNSEKASQDTINSVVEENGEVKELHCKYCKKVGFT 494
503 PT-----FNKFLVLMKAVERKSKADENTPRGSEGSALAGV----- 536
495 HTNMRRHQRIYHERHLFKGYARKGGLEROPRPAQOATONYVVPSTEEBEGEADVV 554
537 -----ADSGATRMOLSKVLTSLPSMALTLNH 563
555 IMDISSNISENLNYITDKIOTNNNTSNCQVIEEMESADLYGICNLLT--PVYVEITON 612
564 LKSTGSPFPYVL--EPGLGASPE----- 585
613 IKTT-QVPVTEDLKREPGLGISTNSEAKKRTTASPPALPKIAETSDDMVPSCSLSPLSI 671
586 -----SKLOOLVE-----KIDROGAVAVASTAGAPPTS 615
672 STTEAVSFHKKEKSVYLSKLLKQLOLQODKLTTPRAGISATEIAKGLPYCVSAPASMLPVTS 731
616 AAPASSASGNOCVTLRLVLSCPRALRLHYGHG--GERPFKCKVCGRAFTRGNL--R 671
732 SRFKRRRTSSP-----SSP-----OHSPALRDFGKPSDGAATDACLTSKK 773
672 AHFVGHKTPA-----ARAONSCPIC-----OKKFT-----NAVTLQOCHVRLHGOI 714
774 SKLESHSDSPWMSLSCDRRETVSPRCFDEYKMSKEWTASSAFSSVCNOQPL-----DL 827
715 PNGGSALSEGGGA-----OENSEQSTASG----- 740
828 SSGVKOKAKGEGKTIVOMESVLDLVYHKKHCSDEKGEKFKESHVQPTCSAVKKRPTTC 887
741 -----PSFPPQOSQOQSPREEEMSEEEED-----EEEDVTD 774
888 MLQKVLNLNEYNGIDLPEVNPADGTRSPCKSLAOPDPDLGPGSGFPAPVYESTPDVCP 947
775 EDSLAGRSSESGE---KAISVRGDESEEVSGAEEEVAT-----SVAAPT 816
948 SSPALQOTSLSSGGLPPLILPIDPSSPPPCPPVLTVAATPPPLPLPTVLPAPSSASAPAP 1007
817 VKEMDSNEKAPDHTLP-----PPPPP----- 838
1008 CSPPLSNMVA-QSPILPILSPVSPSPPIPVPEPLMGAASPGPPTLSSSSSSSSSSSFS 1066

```

```

QY 839 -----DNLDPQPMEGTSDVSGAMEEAKLEGISSPMALTOEG-----EGTSPFLVEELN 890
Db 1067 SSSSSSSPPPLTAISSVSGDNLEASLPMISTFOEILENGLKPRREPQSAEADV 1126
QY 891 LPEAMKRDPGESSGRKACEVCGQSPPTOTALERHOKTHPKDGLPFCVFCRQGLDRTL 950
Db 1127 VQETFNKN-----FVGNVCSPLSIKDLTKHLSTHAEWF-FKCEFCVQLFKDKTDL 1178
QY 951 KKHMLAH 958
Db 1179 SEHREFLH 1186

RESULT 13
US-09-528-706-4
; Sequence 4, Application US/09528706
; Patent No. 6468985
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/528,706
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/516,859
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1776
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-528-706-4

Query Match 4.5%; Score 236; DB 4; Length 1719;
Best Local Similarity 18.1%; Pred. No. 1.6e-07;
Matches 230; Conservative 137; Mismatches 407; Indels 494; Gaps 50;

```

196 DFTSANMRASABPKDEEKPASALQPA-----TLOEVASQEVPELITPA 243
158 PPPPPPPPPGVSGHLNIPILIEELRVLOQROIHOMQTPQICRQVLLLSLGQTVGA 217
244 PAMEQPPEPDERLEAANAQEVNDLGEF-----EEEEEEDEEEEEE----- 282
218 PASPELPTGAASTKPLPLPFSPIKPAOTGKTTTASSSSSSSGAEPKQAFHLYHP 277
283 DDDDELDEDEGEAEASMP-----NENSVKREPE----- 309
278 LGSQHPFVSGVGRSHKPPRPPSPALPGSDQLLASPHLAPPTGTGLAACLGAARGLE 337
310 -----IRDEKP-----EDLLEPXTTSEET-----LE 332
338 --AASPGLLKPNKSGELGYEVISLSEKPGGRHKCRCAKVFSGDSALQIHLRSH--T 393
333 DCEVTPAMQIDR--TKEANGVFEFEMPR-----CONCERKFTTKGLEHMHHT 385
394 GERPYKAVCGNRFTTRGNLKVIFHNRH--EKYPHOMNHPVPEHLDYTTSSGLPYG 450
386 VNMAFKCYCGAFQJIRRRRHRERHEAGLKRKPSQTIQF--SEDL-----ADGKASG 437
451 MSVPPEKAEKEAGTGGGVERKPLVASTALSAATESLTLSTGTSTAVAP-----GL 502
438 ENV--ASKDDSPSPSLGPDCLIMNSEKASQDITINSVVEENGEYKELHPCKYCKKVFET 494
503 PT-----FNKFLVLMKAVEPKSKADEMTPPGSEGSATAGV----- 536
495 HTNMRHORVHERHLIPKGVRRKGLLEFPQPAEQAOQNVYVSPTEPEEGEADVY 554
537 -----ADSGSATRMQLSKVLTSPLSMALLTNH 563
555 IMDISNISEMLNYITDGKIOTNNNTNCDVITEMESASADLYGINCLLT--PVYVEITON 612
564 LKSTGSPPEPYVL--EPLGASPE----- 585
613 IKTT-QYVPTEDLPKEPLIGSTNSEAKKRRTPASPPALPKIKAEIYSDMVPBCSISLPIST 671
586 -----TSKLOOLV-----KIDROGAVAVAVASTAGAPPTS 615
672 STTEAVSFHKERSVYLISKLOLQIOTDKLTPAGISATEIAKLGPCVSPASMLPYTS 731
616 APAPSSASGPNOCVIGLRVLSCPRALRLHYQH--GERPFKVCVGRAFSTRGNL--R 671
732 SRKRRRTSSP-----SSP-----QHSPLAKRFGKPSDKAAWTDAGLTSKK 773
672 AHFVGHKTPA-----ARAONSCPIC-----OKKET-----NAVTLQOHVRLHLSGOI 714
774 SKLESHSDSPAMSLSGRDERETVSPCFDEYKMSKEMTASSAFSSVONQPL-----DL 827
715 PNGSALSEGGA-----QENSSEOSTASG----- 740
828 SSGVKKQAEKGKTPVQWESVLDLSVHKHCSDSEKGEFKESHVSQPTCSAVKKRKPPTTC 887
741 -----PGSFPOPOSOQSPSEEMSEBEED-----EEEEEDVTD 774
888 MLOKVLLNEXNGIDLPEVNPADGTPSPCKSLEAOPDDPLGSGCPAPVETSTPDCP 947
775 EDSLGRGSGSGE--KAISVRGDSSEVSGAEVEAT-----SVAAPT 816
948 SSPALQTPSLSSGQLPPLLPDTPSSPPCPVLTATPPPLPPLPVPLPAPSSASAPHP 1007
817 VKEMDSNEKAPOHTLP-----PPPPP----- 838
1008 CSPSLSNATA-QSPLPLLSPTVSPSPSPIPPEVPLMSAASPGPPTLSSSSSSSSSSSSSFS 1066
839 ----DNLDHPQPMEOGTSDVSGAMEEKEAKLEGISSPMALTOEG-----EGTSPPLVEELN 890
1067 SSSSSSSSPSPPLLSAISSVYSSGDNLEASLPMISFKQELEENGLKPRREPOGAABDDV 1126
891 LPBAMKKDGESSGRKACEVCGSGFPQTQTALEHOKTHPKDQPLFTCVTCROGFLDRATL 950

Db 1127 VOETFNKN-----FVCNVCESPFLSIKDLTKHLSHAEHP-FKCEFCVOLFKDKTDL 1178
QY 951 KKHMLLAH 958
Db 1179 SEHRLILH 1186

RESULT 14

US-09-262-773-4
; Sequence 4, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: Myriad 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 640
; TYPE: PRT
; ORGANISM: human
US-09-262-773-4

Query Match 4.4%; Score 234.5; DB 4; Length 640;
Best Local Similarity 20.2%; Pred. No. 5,7e-08;
Matches 129; Conservative 55; Mismatches 233; Indels 223; Gaps 22;

QY 145 LGAPPLPE--STAPPPPPPPPPGVSGHLNIPILIEELRVLOQROIHOMQMTQIC 202
Db 144 LGAEPSENELODPVQSTPEOSPEETQSPDLGAP--AEQRHQBEELOTLDESE-- 197
QY 203 ROVLLLSLGQTVGAPASPELPGTGAASSTKPLPLPFSPIKPAOTGKTTTASSSSSSSS 262
Db 198 -----VPYBEDP-DLPAERSSGDSSEWVALLTALSQGLVTRKDVAVCFSDQDMS 244
QY 263 GAEPKQAFPHLY-----HPLGQHPFVSGVGRSHKPPRPPSPALPGSTQD 309
Db 245 DLDPQKEFEGYVLEEDCGIVLSFPI--PRPDEISQV-REEEPW-VPDIQEQETQE 300
QY 310 LIASPHL--AFPGTGTGLLAQCLGAA-RGLEAASGLKLP-----KNGSGE 353
Db 301 ----PELISFTYTGGRSKDEDECEQEDLSLEDIHRPVLDPELHQTPDWEIYEDNPGR 356
QY 354 LG--YGEVVISLEK-----PGRRHKCRCAKAVFGSDSALOHLRSHHTGERP 397
Db 357 LNERFGTINISQVNSFVNLRETTPVHPLTGRHNDCSVCGKSPFONSHLVRLHRTHTGKRP 416
QY 398 YKCVGCGRFTTRGNLKVIFHNRKERYPVQANPHVPEHLDYTTSSGLPYGMSVPEK 457
Db 417 YKCEGCKSTYRSSHLARHOKVKNMADP-----KYPLNR 451
QY 458 AEBEAGTPGGGVERKPLVASTALSAATESLTLSTSTAVABGLPTFNKFLMKAVEPK 517
Db 452 KMLEETSPVQAEKTP-----SVEKP 472
QY 518 SKADEMTPPSGSALAGVADSGSATRMQLSKVLVSLPSMA--LTLNHLKSTGSPFP-- 572
Db 473 YRCD-----DCGKHFR-----WTSDLVRHQHTGKPEPCT 504
QY 573 -----PVYLEPLGASPESTSKLOOLVEKIDROGAVAVAVASTAS 609
Db 505 ICKSFSQKSVLTHQRILHGLGKRPVLCGEGDEFSHRY----- 544
QY 610 GAPPTASAPASSASGPNOCVIGLRVLSCPRALRLHYGONGGERPPKCKVCGRAFTSGN 669
Db 545 ----LAHRTTHAAEELIYLCGECGRCFTTHSAFAKHLRSHASVRPCRCNECGKSFNRD- 598

467 KCRYCDVAFVHERYALI-----OHOKSHKNEKRFKCDQCYACHQERHMV 511
Y 644 LHYGOGGERPFCKYCGRAFSTRGNLRAHFVGHKTSPPAARQNSCPICQKFTNAVTLQ 703
512 MKRTHTEKPYACSCDCDFTFRQKOLDMHFKRYHDPNFVPAAFVCSKCGKFTFRNTMA 571
Y 704 OHVRMHLSGQIPNGSALSGGAOENSSQSTASPGSFPOQSPPEEEMSEEE 763
572 RHA-----DNCGLDGGEG---ENGGE--TKKGKRG---RRKMRSKEDSSDSEE 614
Y 764 EDEEEDVTDDESLAGRSGSEGEKALSVRGDSEEVGAEEVATSVAAPTVKEMDSN 823
b 615 NAEPLDDNDEDE-----ETAVEIE-----AEPEVSAEAPAPPSKKRRGR 655
Y 824 EKAPQHTLPPPPPPNDLHPQPM-----EGTSDVSGAMEEAKLEGISSPMALTOE 877
b 656 -----PPCKAATQTKOSQPAALIQVEDONTGEIENIIVEYVK-----EPDAETVEE 701
Y 878 GEGTSTPLVEELN 890
b 702 EEQAQPAVEAPN 714

RESULT 17
CT-US95-08429-5
Sequence 5, Application PC/TUS9508429
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CTCF
NUMBER OF SEQUENCES: 21
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08429
FILING DATE: 15-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/261,680
FILING DATE: 17-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14538A-11-1PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 728 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
CT-US95-08429-5

Query Match 4.3%; Score 228.5; DB 5; Length 728;
Best Local Similarity 21.5%; Pred. No. 1.7e-07;
Matches 119; Conservative 58; Mismatches 221; Indels 155; Gaps 23;

364 EKPGRHRCRPFCAKAVGSDALQIHLRSHTEGRRPKVCNVCGRFTTRGNLKVHFRHREK 423
Y 291 ERP---HCHCLGGRARFTVLLRNHLNLTHTGTRPKCPDCDMAFVTSGLV---RHR-R 342
Y 424 YPHVQMPHVPDEHLDYVLTSSGLPYGMSVPERKAEAEAGTGGVERKPLVASTTSLA 483
343 YKHHEKPFKC-SMCDYA-----SVEVSKLKRHRSTG---ERFQCSLCSYAS 388
Y 484 TESITLLTGSTAVAPGLTFNFKVLMKAVEPRKSADENTPPGSEGSAT---AGVADS 539
389 RDTY-----KLKRMHRTSHGKPYEYGLCHARTQS 419

QY 540 GSATRMOLSKLTVSLPFWALLTNHLKSTGSRFPYV-----LEPLGA-----SPSETS 587
Db 420 GT---MKH-----LQKHTENVAKNPCRPDVTIARKSDGLVRKOHSTIEQG 466
QY 588 KLOQLVEKIDRGAVAVASTASGAPTTSAFAPSSASGPNQ---CVICLRVLSCPALR 643
Db 467 KCRYCDVAFVHERYALI-----OHOKSHKNEKRFKCDQCYACHQERHMV 511
QY 644 LHYGOGGERPFCKYCGRAFSTRGNLRAHFVGHKTSPPAARQNSCPICQKFTNAVTLQ 703
512 MKRTHTEKPYACSCDCDFTFRQKOLDMHFKRYHDPNFVPAAFVCSKCGKFTFRNTMA 571
QY 704 OHVRMHLSGQIPNGSALSGGAOENSSQSTASPGSFPOQSPPEEEMSEEE 763
Db 572 RHA-----DNCGLDGGEG---ENGGE--TKKGKRG---RRKMRSKEDSSDSEE 614
QY 764 EDEEEDVTDDESLAGRSGSEGEKALSVRGDSEEVGAEEVATSVAAPTVKEMDSN 823
Db 615 NAEPLDDNDEDE-----ETAVEIE-----AEPEVSAEAPAPPSKKRRGR 655
QY 824 EKAPQHTLPPPPPPNDLHPQPM-----EGTSDVSGAMEEAKLEGISSPMALTOE 877
Db 656 -----PPCKAATQTKOSQPAALIQVEDONTGEIENIIVEYVK-----EPDAETVEE 701
QY 878 GEGTSTPLVEELN 890
Db 702 EEQAQPAVEAPN 714

RESULT 18
US-08-040-548-1
; Sequence 1, Application US/08040548
; Patent No. 5763209
; GENERAL INFORMATION:
; APPLICANT: Sukhathme, Vikas P.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO THE
; TITLE OF INVENTION: FUNCTIONAL DOMAINS OF DNA BINDING PROTEINS
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: 321 No. 5763209th Clark Street, Suite 800
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/040,548
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coughlin, Daniel F.
; REGISTRATION NUMBER: 36,111
; REFERENCE/DOCKET NUMBER: arc067
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 744-0090
; TELEFAX: (312) 245-4961
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-040-548-1

Query Match 4.3%; Score 226; DB 1; Length 533;
Best Local Similarity 22.1%; Pred. No. 1.7e-07;

```

      REGISTRATION NUMBER: 36,111
      REFERENCE/DOCKET NUMBER: arc067
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: (312) 744-0090
        TELEFAX: (312) 245-4961
      INFORMATION FOR SEQ ID NO: 1:
        SEQUENCE CHARACTERISTICS:
          LENGTH: 533 amino acids
          TYPE: amino acid
          STRANDEDNESS: single
          TOPOLOGY: linear
      MOLECULE TYPE: peptide
      US-08-466-344-1

Query Match              4.3%; Score 226; DB 1; Length 533;
Best Local Similarity   22.1%; Pred.No.1,7e-07;
Matches 124; Conservative 66; Mismatches 190; Indels 180; Gaps 26;

QY 70 PSNSSASAPRPGHSHSQVMDTEHSNPDDSGSGPDPPTWCGERGEESGQFLVAATG 129
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 24 PHSPTMDNYPKLE-----EMLLNSGAQQLCAAGTP-----EGSGGNSSSTS 67
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 130 TAAAGGGGLLAPKLGATPLPESTAPPPPPPPPPPVGVSGCHNLPLELELVLIQQ 189
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 SGGGGGGG-----SNSGSASFNPQGESEQP-----YEHLTTE-SFSIALNNE 110
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 190 RQHQMNTQICROVLLIGQTVGAPASPSELPGTAASSTKPLPLPSPI-----KP 245
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 KAMVEISYPQTR---LPRIYTGRSLERA--PNSGNTLMPEPLFSLVSGLSVMKNP 164
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 246 AQTGTATSSSSSSSSSGBAEPKQAFPHLYHPDLSGHPRSVGVGSHKPT-PAP---- 299
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 PTSSSSAPSPSAASSSSASQSP-----PLSCAVPNDSSPIYSADPTFPPTNDIF 215
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 300 ----SPALPSTQOLLSPHLAFTGTG-----LLAQ-----CLGAA-----RGL 336
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 216 PPQSAPFSPSACTALOYPPRPATKKGFGVOYPMIPDYLFPOOQGDISLTGPDKPOGL 275
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 337 EAAASGGLKP-----KNGSGELGYGEVI--SLEKPG-----                367
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 276 ENRTQGPSLPLSTIKAFATQSGSDLKALNNTYSQLKPSMKRRKYPNRKPTPHERP 335
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 368 -----GRH-----KCRCAKYFGSDSALQILHRSHGTGERPYK 399
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 336 YACPVESCDFRFNRSDDELTRIIRIHTGQKPFQCRCIMRNFRSRDLTHTRTHGEKPEFA 395
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 400 CAVCGKRFT---TRGLKLVHFHHRER-----YRHVQNPHVPHEILD 439
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 396 CDICGKKFARSDERKRRTKIHLRQDKRADKSIVASPAASSLSISSPVATSYSPATTS 455
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 440 Y---VTSSGLPLGMVSYPEKAEEAGTGGGVERRPLVAST---TALSATESL 487
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 456 FESPVTYTISSP-GSSITYPSRA--HSGFP-----SPVAITTFASVPAPAEPTYOVSSPSFA 506
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 488 TELLT-STSTAVALGPLTFN 506
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 GVSSSFSTSTGLSDMTATFS 526
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 20
5206152-2
: Patent No. 5206152
: APPLICANT: SUKHATME,VIKAS P.
: TITLE OF INVENTION: CLONING AND EXPRESSION OF EARLY GROWTH
: REGULATORY PROTEIN GENES
: NUMBER OF SEQUENCES: 18
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/249,584
: FILING DATE: 26-SEP-1988
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 179,587
: FILING DATE: 08-APR-1988
: SEQ ID NO:2:

```

LENGTH: 533

Query Match 4.3%; Score 226; DB 6; Length 533;

Best Local Similarity 22.1%; Pred. No. 1.7e-07; Mismatches 124; Conservative 66; Indels 180; Gaps 26;

70 PNSSASAPREPSHSQVMDTEHNSPPDSSGSGPPDPWGERGEESGQFLVATG 129
 24 PHSPTMDNYKLE-----EMLLSNGAPQFLGAGTP-----EGSGGSSSSS 67
 130 TAAAGGGGGLILASPKLGTPLPESTPAPPPPPPPPGVSGHLNIPILLEIRVQQ 189
 68 SGGGGGGG-----SNSGSAFNPQCEPSEOP-----YEHLTTE-SFSDIALNE 110
 190 RQHHOMQMTFOICQOVLLGSLGQVVGAPASPSLPGTGAASRKLPLPLPSPI----KP 245
 111 KANVEVSYPQOTR---LPIITYGRSLEPA--PNSGNTLMPEPLFSLVSGLSMTNP 164
 246 AOTGKTATASSSSSSSGAEPKQAFPHLYHPLGSOHPFSVGGRSHKPT-PAP----- 299
 165 PRSSSAPSPASASSSSASOSP-----PLSCAVPSNDSSPTYSAPPTPTNTDIF 215
 300 ---SPALPGSTDLIASPHLAFPGTTG-----LLAQ-----CLGAA-----RGL 336
 216 PEPQSOAFPSAGTALQYPPPAYPATKGGFYVMPIDYLPQQGDLISLGPDKPFOGL 275
 337 EAAASGLKLP-----KNGSELGGEVI--SLEKPG----- 367
 276 ENRTOQPLPLSTIKAFATQSGODLALANTTYOSQLKPSRMKYPNRPDKTPPHERP 335
 368 -----GRH-----KRCFCAYVGSALQIHLRSHTEGEPYK 399
 336 YACPVESCDRFSDELTFRIRIRITGOKPPOCRICMRNRSRSHLTHITHGEKPPA 395
 400 CAVCGNFT---TRGNLKVHFRHREK-----YPHVQMPHPVEHL 439
 396 CQICGRKFAASDERKRRHTKILRQKDKADKSVASPAASLSYSPVATSPAPATTS 455
 440 Y---VTSSSLPGMSVPEKAEAGTGGVEKRPVAST-----TALSATESL 487
 456 FSPVPTYSVP--GSSYTPSPA--HSGFP-----SPVATTPASVPAPPTQVSSFPSPA 506
 488 TLIST-GTSTAVALPGLPTFN 506
 507 GVSSSFSTSTGLSMTATFS 526

RESULT 21
 US-08-061-376-5
 Sequence 5, Application US/08061376
 Patent No. 6175000

GENERAL INFORMATION:
 APPLICANT: Evans, Glen A.
 APPLICANT: Djabali, Malek
 APPLICANT: Seller, Licia
 APPLICANT: Parry, Pauline
 TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23
 TITLE OF INVENTION: TRANSLOCATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
 STREET: 444 South Flower Street, Suite 2000
 CITY: Los Angeles
 STATE: California
 COUNTRY: USA
 ZIP: 90071

COMPUTER READABLE FORM:
 MEDIUM TYPE: IBM PC floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/061,376
 FILING DATE: 13-MAY-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Reiter, Stephen E.
 REGISTRATION NUMBER: 31,192
 REFERENCE/DOCKET NUMBER: P41 9387
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619)546-4737
 TELEFAX: (619)546-9392
 INFORMATION FOR SEQ. ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3969 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-061-376-5

Query Match 4.3%; Score 226; DB 4; Length 3969;

Best Local Similarity 19.9%; Pred. No. 2.3e-06; Mismatches 245; Conservative 117; Indels 394; Gaps 52;

2 AOTGSSSRILGPGCGEPAERGDASEEHNPQYCAKCAQFSDPTEFLAHONSCCTD-PPV 60
 452 APSGSSSEKSSAASQHSQMSDSSSSPSVDSTDSQASERIQYLPERSDPEVHP 511
 61 MYLIGQENPSNSASS---APREGRSRQVMDTEHNSPPDSSGSGPPDPWGERRG 116
 512 LPIQSPEENESNDRRSRYSVESFSGRTTKLSTLQSAPOOQTSSPPPLPTPPPL 571
 117 EESS-----GQFLVATGTAAAGGGGLILASPKLGTPLPESTPAP 159
 572 QPASSISDHTPMLPPIPLASPLPASTAPMOGRKS--ILKEPTRTWISL-KHSRSE 629
 160 -----PPPPPPPPPGVSG----- 173
 630 YPSAKYAKEGLIRKPIFDNFRPPPLTPEDYGFASGFSASGTAAARLFGPLHSGTRFDM 689
 174 HLNIPILERLYLOQROIHMOMTEQICROYLLGSLQTYGAPAPSPSLPGGAASS- 232
 690 HRSPL-----LRAPRTPEAHSRIFEVTL-----PSRRTSAGTSSSG 729
 233 ---TKPLPLESPIK---PAOTGKTATASSSSSSSGAEPKQAFPH---YHPIGS- 280
 730 VSNRRKRVFSPTRSPRBSHMRTRSGRLSSSELSPITPSSVSSLSISVPLATS 789
 281 -----QHPPSVGVG---RSHKPTPAPSPALPGSTDLIASPHLAFPGTTGIL 326
 790 ALNPFTFPSSHLSLQSGESAENKQRPKOTSAFAPFPSSS-----SPRPLFPWFTP--G 841
 327 AOCGLAARGLAASAPLL----- 353
 842 SQ---TERGRNKKDAPBELSKDRDADKSVKDKSHERDEREKREKREKRRKKS-- 896
 354 LGYGEV-LSLEKPGGHHKRCFAKVFSDSALQIHLRSHTEGEPYKAVCGNFTT--- 409
 897 ---ETQSSALYPGVYSKE---KYVGEDVATSSAKATGRKSSSHSGDITIVTL 949
 410 -----RGNLKVHFRHREKYPHVQMPHPVEHLIDYVTTSSGLPYGMSVP 454
 950 GDTTAVAKTKILIKKGNL----- 968
 455 PEKAEFEAGTPGGGVEKPLVASTALATESLTLTLSTGTAVALPGLPTFNKFV---LM 511
 969 -EKTNDLGPPTAPSLERKTLCLSTPSSSTYVKHSTSSISGMSLAQADKLPTDTRVVASL 1027
 512 KA-----VEPKSKADENTPPGSEGSALAGVADSGSATPMOLSKLVTSLPSMALTLTHL 565
 1028 KAKAOLCKIEKSKSLQTDOPKAGOE---SDS-SETSVRGRIRHVCRAAVALGRKR 1082
 566 STGSEPPFVLEPLGASP-SETSKILOOLVERKIDROGAVAVAVASGAPPTTAPAPSSSAS 624

Db 1083 AV---FPDDMPTLSALPWEREREKILSSMGNDKSSIASSEDAEPLAPIKPIKPYTRNK 1138
OY 625 GNOCVI-----CLRVLSCPRALRHGQHGGERPKVCVCGRAFSTRGLRAHFVGHKT 679
Db 1139 APOEPVKKGRRRRCGQCP-----CCQVPEDCGVCTNCLD-----KKEFGGRNT 1183
OY 680 SPARAONSCPIQ-----KKFTNAVTLQAHVRLHGLQIPNGSALSSEGGAAQEN 731
Db 1184 KKOCCMRKCOQLQWMPKSKAYLQKQAKAVKKKK-----KSKTSE-----KRD 1226
OY 732 SEEOASTASC---PGSFPOPOSOQ--PSPEEEMSEEEDEEEDVTDDESLAGSSEGG 787
Db 1227 SKESVYVKNVDSOSKPTPSAREDPAPKSSSEPPRPKRKVEEK--SEEGNVASAPGESKQ 1284
OY 788 EKAISVRGSEEVSGAEEVATSVAPPTVKEMDSENAKAPQTLPPR---PPRDLDH 843
Db 1285 ATTPARKSKSKQVS--QPALVTPPOPTT---GPRKREVPKTPSPKKOPPPES--G 1337
OY 844 POPMEGTSVDSGAMEEBAKLEGISSPMALTOEGEGTSTPLVEELNLEPAMKKDPGE-- 901
Db 1338 PQOSQOKKVAAPRSIPYKQPKKEKEKPPVKNQENAGTLNLTSLNSNGSSKQKIPADGV 1397
OY 902 -----SSGRKA-----CEVCGSGFPPTQ 918
Db 1398 HRIIVDFKEDCEAENWEMGGLITLTVPTPRVVCFLCASSGHVEFYVCQVCEPEF-HK 1456
OY 919 TALEHQHTHPKQGRPLT-----CVFCQGFDRATLKHHMLAHHQVPPFAHGPONIA 973
Db 1457 FLELENER--PLEDOLENNCCRRCKFC-----HVCGRHO-----A 1490
OY 974 TLSLVP--GCCSSISPGLSP-FPRKDDPT 1000
Db 1491 TKOLLECNKCRNSYHRECLGPNYPTK--PT 1518
RESULT 22
US-08-475-844-9
Sequence 9, Application US/08475844
Patent No. 5972643
GENERAL INFORMATION:
APPLICANT: Lobanenkoy, Victor V.
APPLICANT: Nelman, Paul E.
APPLICANT: Klenova, Elena M.
APPLICANT: Goodwin, Graham H.
APPLICANT: Filippova, Galina N.
APPLICANT: Collins, Steven J.
TITLE OF INVENTION: CTCF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend Kourlie and Crew
STREET: One Market Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,844
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/261,680
FILING DATE: 17-JUN-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14538A-11-1
TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 727 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-475-844-9
Query Match 4.3%; Score 225.5; DB 2; Length 727;
Best Local Similarity 20.7%; Pred. No. 2.7e-07;
Matches 115; Conservative 61; Mismatches 207; Indels 173; Gaps 23;
OY 364 EKGGGRKRCACAVFGSDSLQIHLSHTGTERPYKCNVCGNFTTRGNLKVHFRREK 423
Db 291 ERP--HKCHLGCAGFAFVTLRLNHLNTHGTGRPHKCPDCDMAFVYSGELV---RRR-R 342
OY 424 YPHVQMPHPVPEHLDYVITSGLPYGMSVPPEKAEAEAGTPEGGVKRPVLVASTALSA 483
Db 343 YKHTHEKPKC-SMCDYA-----SVEYSKLKRHRSHTG---ERFQCSLCSTAS 388
OY 484 TESLTLTSTGTSTAVAPGLPTFNKFLMKAVBPKSKADENTPPGSEGSAT---AGVADS 539
Db 389 RDTY-----KLRHMRTHSGEKPEYECYICHARFTQS 419
OY 540 GSATRMQLSKLVTLSPSMALLTNHLKSTGSPFPYV-----LEPLGA-----SPSETS 587
Db 420 GT--MKMH-----ILQKHTENVAKFHCPCPHCDTVIARKSDGLVHLRKQHSYIEQG 466
OY 588 KLOQIVKIDROGAAVAVASTASGAPTSAPAPSSASGPNQ---CVICLRVLSCPRALR 643
Db 467 KRCRYCDVAVERKALY-----QHOKSHNEKFKCDQCYACQOEKMT 511
OY 644 LHYGQHGGERPKCKVCGRAFSTRGNLRAHFVGHKTSAPARAONSCPIQCKKFTNAVTLQ 703
Db 512 MHRTHTEGERPYACSHCDKTFROKQLDLHMFKRYHDNPVPAFVCSKCGKTFRRNTMA 571
OY 704 QHYRMHGGQIPNGSALSSEGGAAQENSSSEOSTASGPSFPQPOQOOPSPPEEMSEEE 763
Db 572 RHADNCAG--PDG---VEG---ENGGETKKSIRGRRKRRKSKEDSSDNAEPDL 619
OY 764 EDEEEEDVTDDESLAGSSESGEKAISVRGDSSEVSGAEEEVATSVAPPTVKEMDSN 823
Db 620 DQWDEDE-----EPAVEIEPPE----- 637
OY 824 EKAPQHTLPPPP-----PPDNIDHPQPMQGTSDV-----SGAMEE--EAKLEGI 867
Db 638 ---POVTPAPPAKRRGRBPGRTNPK-QNOPALIQVEDQNTGAIENIIVEVKKRPD 693
OY 868 SSPMALTOEGEGTST 883
Db 694 ABPAEGEEEAQAPAT 709
RESULT 23
PCT-US95-08429-9
Sequence 9, Application PC/TUS9508429
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CTCF
NUMBER OF SEQUENCES: 21
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08429
FILING DATE: 15-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/261,680


```

1087 TPPAATTTTATTATTAVP-----APTPAP----- 1115
304 PGSTDOLIASPHLAPRTTGLLAOCLGARCLEAASPLIKRKNCSGE-LGYEVISS 362
1116 -----ORLILSPD-----MOARLPSGEVVISIOLASL 1142
363 LEKPGGHKRCFKACVFGSDSALQILRSHTEGPRYKCNVGNFTTGG-----NLKVHRH 418
1143 AORP-----VANAGGSKPLTFQIOGNKLTLLTGAOYROLAVGCP 1180
419 RHREKYPHVQMN-----PAPVPEHLDYVITSSGLPYGMSVVP 455
1181 RPLQMPRTYMNNTGVVVKLVVROAPRDGLTPVREPLAPAPR-----PSSGDPAILNPPR 1233
456 EKAEDEEGTGGGVCKRLVASTALISATESLTLSTGCTSAVAPGLPTFNKFLMKAVE 515
1234 TLPPGLRPLPLGTARADY-----PTPLVVR-----LTKLVH 1266
516 PKSKADENTPFGSE-----GSAIAGVADS-----GSATRMOLS----- 548
1267 SPSPREVSASAPGARPLTISPLHVPSSLPFGPASSPMPFIPNSSPLASPVSTYVSPILSSSL 1326
549 --KLVTSLPS-----WALLTNHLKSTG---SPFPYVLEPLGAS 582
1327 PISVPTPLPAPASAPLTIPIAPLTVSASGAPALLTSVTPPLAPVAPAPGPPSLQPSGAS 1386
583 PSSTSKIOQLVEKIDROGAVAVASTASGAPTTAP---APSSS-ASGPNOCY--ICLRVL 636
1387 PSASALATLGL-----ATAPLSSSQTPGHPRLALAPLPSHVPGINSTVAPACSPVL 1436
637 SCPRALRHYGONGGERPFCVKVCGRAFSTRGNLRAHFVGHKTSPPARAONSCPIQKKF 696
1437 VPASAL-----ASPPSAPNAPR----- 1455
697 TNAVTLQOHVMIHGGQIPNGSALSSEGGAAOENSSEOSTASGSPGSPPOGQSPREE 756
1456 -----QASLLAPASSAQ---ALATPLAPMAQPTAILLAPSPAP 1491
757 EMEDEEEDDEEEDVTDSDLAGRSESGEKAISVRGDESEVSGAEVATSA----- 812
1492 PTLA-----PPLVLAAPSPGAPVLAASSQTPVPVWAPASSTPCTSLASASP 1534
813 --APTVKEMDSNEKAPONTLPPRRPPDNLDPQMEQGTSDVSGAMEEAKLEGISSP 870
1535 VPAPTYPLVAPBST---QTMPLAPVPSR---LPSPASTQTLALAPAL---APTLGGSSP 1583
871 MAALTOGEGISTPL---VEELNLEPEA-----MKDPEBSSGRKACEVCGQSTP 916
1584 SQTLSL---GTGNPQGPRTPTOTLSLTPASSLVPTPAOTLSLAPGPPILG-----P 1629
917 TOTALBEOHKHNPDKGRLFTVCFCRQGFIDRATLLKKHMLLAHQVPRPAPH---GPQNTA 973
1630 TQT-----LSLAAPPLAPASPVGAPAPAH 1653
974 TILSLVPGCS-----SIPSPGLSPFP 994
1654 TILFLAPASSASLAPASVQTLTLSPAP 1681

```

RESULT 27
 JS-08-102-942A-4
 Sequence 4, Application US/08102942A
 Patent No. 5726288

GENERAL INFORMATION:

APPLICANT: Call, Katherine M.
 APPLICANT: Glaser, Thomas M.
 APPLICANT: Ito, Caryn Y.
 APPLICANT: Buckler, Alan J.
 APPLICANT: Pelletier, Jerry
 APPLICANT: Haber, Daniel A.
 APPLICANT: Rose, Elise A.
 APPLICANT: Housman, David E.
 APPLICANT: Bruening, Wendy

```

; APPLICANT: Darveau, Andre
; TITLE OF INVENTION: Localization and Characterization of the
; TITLE OF INVENTION: Wilms' Tumor Gene
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/102, 942A
; FILING DATE: 02-AUG-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-5194A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-102-942A-4

```

Query Match 4.2%; Score 222; DB 1; Length 449;
 Best Local Similarity 25.6%; Pred. No. 2.5e-07;
 Matches 100; Conservative 38; Mismatches 135; Indels 118; Gaps 20;

```

QY 124 LVNATGTAAGGGGLIILASPKLGA---TPV-----PPEST-----PAPPPPPPPPPPG 169
DB 11 LPPAVSSLLGGGGGCGGL--PVSGARQMAPVLDPAFGASAYGSLGGPAPPPAPPPPPPP 67
QY 170 VCSGHNIPLILEELRYVQOQRIHOMQTEQICRGVLLGSLGQ---TVGA-----PA 219
DB 68 -----PHSFIKQEPSWGGAEPRHEEQCLSAFTLHFSQFTGTAGCRGPFPP 115
QY 220 SPSEL-PTGAASSTKPLPLPSPIKPA--GTGKTTA-----SSSSSSSSGAEPPKQA 270
DB 116 PPSQASSGQARMFPNAPVLPSCLESQPTIRNOGYSTVTFDGAESYGHTPSHHAQFPNHS 175
QY 271 FPHLYHPILGSOHPFSVGVGSRSHKPTP-----APSPALPGSTDOLIASPHLA---PPG 320
DB 176 FKH-EDPMGQD---GSLGEOQYSVPPPVYGCNTPTDSTGSOALLIRTPYSSDNLTYOM 229
QY 321 TTGLLAQC-----LGAARGLBAASPLGPKNGSGELGYGEVVISLEKP---GGRH 370
DB 230 TSGL---ECMTWQMNIGATLKGMAAGSSSVKWTGQGSNHGTYESENHTAPILCGAQY 286
QY 371 K-----CRF---CAKVFSGSDALQIHRS 391
DB 287 RIHTHGVFRGIQDVRRVSGVAPTLVRSASETSEKRPFMCAVPGCKNRYFKLSHLOMHSRK 346
QY 392 HNGERPCKNV--CGNRTTGNLKVHFRH 420
DB 347 HNGEPYQCDPKDCERRRSRSDQLKRHORH 377

```

RESULT 28
 US-09-037-179B-4
 Sequence 4, Application US/09037179B
 Patent No. 6316599
 GENERAL INFORMATION:

APPLICANT: Call, Katherine M.
APPLICANT: Glaser, Thomas M.
APPLICANT: Ito, Caryn Y.
APPLICANT: Buckler, Alan J.
APPLICANT: Pelletier, Jerry
APPLICANT: Haber, Daniel A.
APPLICANT: Rose, Elise A.
APPLICANT: Hausman, David E.
APPLICANT: Bruening, Wendy
APPLICANT: Darveau, Andre
TITLE OF INVENTION: Localization and Characterization of the
FILE OF INVENTION: Wilms' Tumor Gene
FILE REFERENCE: 0050.1312-011
CURRENT APPLICATION NUMBER: US/09/037,179B
CURRENT FILING DATE: 1998-03-09
PRIOR APPLICATION NUMBER: US 08/102,942
PRIOR FILING DATE: 1993-08-02
PRIOR APPLICATION NUMBER: US 07/614,161
PRIOR FILING DATE: 1990-11-13
PRIOR APPLICATION NUMBER: US 07/435,780
PRIOR FILING DATE: 1989-11-13
PRIOR APPLICATION NUMBER: US 07/795,323
PRIOR FILING DATE: 1994-09-27
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 4
LENGTH: 449
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Murine
US-09-037-179B-4

Query Match 4.2%; Score 222; DB 4; Length 449;
Best Local Similarity 25.6%; Pred. No. 2.5e-07;
Matches 100; Conservative 38; Mismatches 135; Indels 118; Gaps 20;

124 LVATGTAAGGGGGGILASPKLGA--TPL-----PPEST-----PAPPPPPPPPP 169
11 LPPVSSLGSGGGGCGL--PVGARQWAPVLDFAFPGASAGSLGPPAPPPPPPPPP 67
170 VGSGLNPLLEELRVLDQROQHOMQTEQICROYLLGSLGQ--TVGA-----PA 219
68 -----PHSFIKQEPKSWGAEPNEQCLSAFTLHFSQFTGAGACRGYGPFGPP 115
220 SPSEL-PGTGAASSTKPLPLPFSPIKA--QTGKTA-----SSSSSSSSGAEPPKQA 270
116 PPSQASGQARMEFPAAPYPLPSCLESQPTIRNOGISTVTFDGAFTYGHTRPSHHAAQFPNHS 175
271 FFHLVHPLGSHPFPSVGVGRSHKPTP-----APSPALPGSTDLQILASPHLA--FPG 320
176 FKH-EDPWGQQ-----GSLGEQOYSVRPVYGGCHTPDSCGSQLALLRTYSSDNLYGM 229
321 TTGLLAQC-----LGAARGLAASPELLKPKNGSGELGYEVSSLEKP---GGRH 370
230 TSOL---ECMTWQNLGATLTKGMAAGSSSVKWTGQSNHGTGYESENHTAPILCGAQY 286
371 K-----CRP--CAKVEGSAIQIHLRS 391
287 RIHTHGVREGIADVARYSVAPTLVNASASETSEKRRPMAICPGCNKRTFKLSHLOMHSRK 346
392 HTGRRPYKCNV--CGNRFTTTRGNLKVHFNRH 420
347 HTGKPYOCDFKDCERRFSRSDQLKRNORRH 377

RESULT 29
US-08-224-482-2
Sequence 2, Application US/08224482
Patent No. 5837692
GENERAL INFORMATION:
APPLICANT: Mercola, Dan
APPLICANT: Adamson, Eileen D.

TITLE OF INVENTION: Inhibition of the Mitogenic Activity of
TITLE OF INVENTION: PDGF by Mammalian EGR
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRES:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/224,482
FILING DATE: 07-Apr-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-ME 9913
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-224-482-2

Query Match 4.2%; Score 220.5; DB 2; Length 496;
Best Local Similarity 22.5%; Pred. No. 3.6e-07;
Matches 120; Conservative 59; Mismatches 179; Indels 175; Gaps 25;

97 PPDSSGSPDDPTWGPBRGEESSGQFLVAATGTAAGGGGILASPKIATPLPESTP 156
9 PQLFGAAGTP-----EGSGGNSSTSSGSGGGGG-----SNGSSAFNNGPEP 52
157 APPPPPPPPPEVSGSHLNIPLLEELRVLDQROQHOMQTEQICROYLLGSLGQYTG 216
53 SHQP-----YELTTE--SFSDIALNNEKAMETSYPSQYTR---LPPIYTGK 96
217 APASPELPGTGAASSTKPLPLPFSPI-----KPAOTGKTASSSSSSSSGAEPPQAFF 272
97 FSLERA--PNSGNTLMPEPLFSLVSGIVSMNTNPTSSSAPSPASSSSSASQSP----- 149
273 HLVHPLGSHPFPSVGVGRSHKPTP-PAP-----SPALPGSTDLQILASPHLAEPGTT 322
150 ---PLSCAVPNSDSSPIYSAAFTFPPTNDIFPEPQSAFPGSAGTALQYPPPAVPATK 205
323 G-----LLAAQ-----CLGAA-----RGLEAASPGLLKP-----KNGSGL 354
206 GGFQVPMIPDYLPFPOQOGLSLSTPDKPFGLENRTOQPSLTPLSTIKAFATQSGSODL 265
355 GYGEVI--SLEKPG-----GRH----- 370
266 KALNTTYQSOLIKPSMRKYVPNRPKTPRHERPYACVESCDRRSRDELTRHRIHTG 325
371 ---KCRCAKVFPGSALQIHLRSHTGERPYKCNVCGNRFT---TRGNLKVHFNRHRE 422
326 QKPFQCRICMRNFSRSDHLLTTHRTGKRPFCADICGRKPARSDERKRNHTKILHROKCK 385
423 K-----YPHVQNPMPVREHLDY---VITSSGLPYGMSVPEKAELEAG 463
386 KADKSVASPAASLSLSYSPVATSTPSPATTSPPSPVPSTSYSSP--GSSTYSPSP--HSG 442
464 TPGGGERKPLVAST-----TALSATESLTLST--GTSTAVAPQLPTFN 506
443 FP-----SPSVATTFASVPAPFPTQVSSPSPSAGVSSSFTSTGLSDMTATFS 489

Db 239 NOMNLGATLKSHSTGYESDNHTTPIILCGAQRIRHTHGVFRGIQDVRVPGVAPTLVRSAS 298
QY 350 GSGELGYGEVSISSLEKRGHRCRF--CAKVFSGDSALQIHLRSHTGERPKCNV--CGN 405
Db 299 ETSE-----KRP---FMCAVPGCNKRYFKLSHLOMSRKHTGERPKQCFDPCR 345
QY 406 RFTTRGNLKVHFHRH 420
Db 346 RFSRSDQLKRHRH 360

RESULT 32

PCT-US95-05523-4
Sequence 4, Application EC/TUS9505523
GENERAL INFORMATION:
APPLICANT: The Wistar Institute of Anatomy and Biology
TITLE OF INVENTION: M1 Monoclonal Antibodies and
TITLE OF INVENTION: Methods of Use Therefor
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/234,783
FILING DATE: 28-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST488PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-05523-4

Query Match 4.1%; Score 217; DB 5; Length 429;
Best Local Similarity 25.1%; Pred. No. 5.1e-07;
Matches 94; Conservative 38; Mismatches 123; Indels 120; Gaps 18;

QY 133 GGGGGLILASPKLGATP-----LPPEST-----PAPPPPPPPPPPGVSGSHLNP 178
Db 19 GGGGGCAL--PVSGAOMAPV--DFAPRGSATVSLGSPAPPPAPPPPPPPPHS----- 70
QY 179 LILBELKVLQROIHQMOMTEICROYLLGLSGQ--TWGA-----PASPSEL-PGT 227
Db 71 -----FKQKPSWAGAEHNEOCISAFIVHFSGQFTGACACRYGPPPPPSQASSGQ 124
QY 228 GAASSTKRLPLPFSIKPA-----OTGKTASSSSSSSSGAEPKQAFHLYHPLG 279
Db 125 ARMPNAPYLSCLESQRAIRNOGVSTVFDGTPSYGHTPSHHAQEPNHSFKR-EDPVG 183
QY 280 SQAPFSVGVGRSHKPTP-----APSPALPGSTQDLIASPHLA--FPSTTGL----- 324
Db 184 QQ-----GSLGEQGVSVPPVYGCHPTDSCSTGSQALLRTPYSSDNLYQMTSQLECMTW 238

QY 325 ----LAAQCLGARGLEA-----AASPGLKPKN 349
Db 239 NOMNLGATLKSHSTGYESDNHTTPIILCGAQRIRHTHGVFRGIQDVRVPGVAPTLVRSAS 298
QY 350 GSGELGYGEVSISSLEKRGHRCRF--CAKVFSGDSALQIHLRSHTGERPKCNV--CGN 405
Db 299 ETSE-----KRP---FMCAVPGCNKRYFKLSHLOMSRKHTGERPKQCFDPCR 345
QY 406 RFTTRGNLKVHFHRH 420
Db 346 RFSRSDQLKRHRH 360

RESULT 33

US-08-246-489-2
Sequence 2, Application US/08246489
Patent No. 6225049
GENERAL INFORMATION:
APPLICANT: Ian, Michael S.
TITLE OF INVENTION: NOVEL HUMAN INSULINOMA-ASSOCIATED CDNA
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Knobb, Martens, Olson & Bear
STREET: 620 Newport Center Drive
CITY: Newport Beach
STATE: California
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/246,489
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/901,715
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH012.012A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-0176
TELEFAX: (619) 235-8550
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 510 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-246-489-2

Query Match 4.1%; Score 216.5; DB 4; Length 510;
Best Local Similarity 22.7%; Pred. No. 6.9e-07;
Matches 140; Conservative 46; Mismatches 229; Indels 201; Gaps 30;

QY 133 GG--GGGILASPKLG--ATPLPESTPAPPPPPPP-----PPPP 168
Db 22 GGEDDRLALISPGCGARPPAPSPVPPPLPPPPAERAAHAAALAACAPGPQPPQ 81
QY 169 GVGSGHLN-----IPLILELRLVLOQROIHQV-----QMTPOICROYLLIGSL 211
Db 82 GPRAAHFGPRAAHAPRPLSPTRPVSGREHEKHKYFERSFNLSYSAISFPRAALLGG 141
QY 212 GQYGAAPRSPSELPGTAASSTKPLPLPFSPIKPAQGTGTTASSSSSSSSGAE-----P 266


```

: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 449
: TYPE: PR1
: ORGANISM: Homo sapien
US-09-037-179B-6

Query Match      4.1%; Score 215.5; DB 4; Length 449;
Best Local Similarity 25.4%; Pred. No. 6.9e-07;
Matches 97; Conservative 35; Mismatches 133; Indels 117; Gaps 18;

QY 133 GGGGGLIASPKIGATP-----LPEST-----PAPPPPPPPPGVSGHLNIP 178
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 19 GGGGGCAL--PVSQAQMAPVLDPAFPGASAYGSLGFAFAPPPPPPPPPH----- 70
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 179 LLELRVLQQRQIHOMQTEQICROYLLGSLQ--TVGA-----PASSEL-PGT 227
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 71 -----FIKQEPSWGAEPHEEOCLSAFTVHFSGFTGACGRYPGPPPSQAASSG 124
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 228 GAASSTKLLPLFSPKIPK-----QTGKTTASSSSSSSSGAPPKQAFLLYHPLG 279
      : | : | | | | | | | | | | | | | | | | | | | | | | | | |
DB 125 ARMPNAPYLPSCLESQPALRNOGYSTVTFDGTBSYGHTPSHHAAPFNHSFKH-EDPMG 183
      : | : | | | | | | | | | | | | | | | | | | | | | | | | |
QY 280 SQHPFSVGVGGRSKRPT-----APSPALPGSTQDLIASPILA--PFGTTGLAAQC 329
      : | : | | | | | | | | | | | | | | | | | | | | | | | | |
DB 184 QQ-----GSLGQOQYVSPRPVYGCHPTDCTGSQLLLRTPYSSDNLQYMTSQL--EC 235
      : | : | | | | | | | | | | | | | | | | | | | | | | | | |
QY 330 -----LGAAGCLEAASPGCLKPKNGSGELGYEIVISLEP-----GGRHK----- 371
      : | : | | | | | | | | | | | | | | | | | | | | | | | | |
DB 236 MTWNQNLGATLKGVAAGSSSVKWTBQSNHSTYESDNTHTTILCGAORYRHTHGVFR 295
      : | : | | | | | | | | | | | | | | | | | | | | | | | | |
QY 372 -----CRF--CAKYGPSALQIHLRSHTGERPYKC 400
      : | : | | | | | | | | | | | | | | | | | | | | | | | | |
DB 296 GIODVRNPGVAPTLVRSASETSEKRPEMCAYPGCNKRYFKLSLDMHMSKHTGEKPYQC 355
      : | : | | | | | | | | | | | | | | | | | | | | | | | | |
QY 401 NV--CGNRTTNGNLKVHFHRH 420
      : | : | | | | | | | | | | | | | | | | | | | | | | | | |
DB 356 DFKDCERRRSRSDQLKRHORRH 377
      : | : | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 36
US-08-545-860D-55
: Sequence 55, Application US/08545860D
: Patent No. 6040140
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: APPLICANT: Canaan, Eli
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
: TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
: NUMBER OF SEQUENCES: 94
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
: ADDRESSEE: No. 6040140r1s
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/545, 860D
: FILING DATE: 07-MAR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04496
: FILING DATE: 22-APR-1994
: PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: PCT/US92/10930
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/327, 392
FILING DATE: 19-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/320, 559
FILING DATE: 11-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062, 443
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971, 094
FILING DATE: 30-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888, 839
FILING DATE: 27-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805, 093
FILING DATE: 11-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TUU-1262
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 35
SEQUENCE CHARACTERISTICS:
LENGTH: 1093 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-545-860D-55

Query Match 4.0%, Score 211.5, DB 3; Length 1093;
Best Local Similarity 20.6%, Pred. No. 4.1e-06;
Matches 192; Conservative 88; Mismatches 321; Indels 329; Gaps 41.

QY 3 QEIGSSSRRLGPGCEPARGGDASDEHHPOVQAKCAOFSDPEEFLAHONSCTDPPVMV 62
DB 289 QETSSESS-----ESKGRKSSSHLSHKCKLSSGKGVSSFTASSSSSSSSSS 338
QY 63 IIGQENPNSNSASSAPRPEGHSR-SQVMDTEHSNPPDSSGSGPPDPT----- 109
DB 339 -GGPFGPAVSSLOSSPDSAPPKLEQPEEDKYSKPTAPAPASAPPSAPRPAADLEQ 396
QY 110 -----WGPERK-----GESSSGOFLVAATG-----TAAGGGGLLASPLKATPLPE 153
DB 397 KVFVSGFGPIRFTSTTSSSGRARAPSDPDYKSPHYTGSGASACTHKMPALSKATPVAD 456
QY 154 STP-----APPP-----PPPPPPG----- 169
DB 457 ETPEPTGLKKHKHKKSRSHGPRGRGSRNKEGTGCPAPASLPQAOLAGTTAASPESG 516
QY 170 ---VSG-----HLNPLILEELRYLQORQI-HQOMQTEQICROYLL 207
DB 517 GSIVSSSGIGLSSRTFGPSGLPSLSLEBPPLLAGITYSNKPDISSHSGMRAVCSPLS 576
QY 208 LGSIGQWVGAPASP-----SELPGTGAASSTKPLLPFSPYKPAQTGKTTASSSSSS 259
DB 577 SSLIGPP-GTSALPLRSRPFSTLPSSSSASISTTOVFSL-----ASTFSL 622
QY 260 SSSGAEPKQAFHLY-HPLGSOHPSPVSGVSGSHKP-----TPAP 299
DB 623 PST-----HFFGTGPAVNPPLSQAESSTHTEPDLDCSFRCRGTSPOESLSSMP 672
QY 300 SPALPGSTDLIASP-----HLAPFGTT-----GLIAQACIGANGLEA- 338
DB 673 ISSLPALFDOTASACGGGOLDPRAAGCTTMEQDLLEKQDGEAGVNIYEMLKALHALQKE 732

```

```

OY 339 -----AASPLKPKNG-----SGELGYEVIS 361
Db 733 NORLQEQILSTAKKERIQLINVLVSFPPALPALPANGPVPGRYGLPPQAGSSDLS 792
OY 362 SLEKPGGRHKCRCAKVFSGDSALQIHLRSHTGERPYK-CNVCGNRFTTNGNLKVHFRH 420
Db 793 TSKSPGKSS-----LGLDNL-----STSEDPHSGC-----PSRSSSSLSFHS 833
OY 421 REKYPHVOMNHPVREHLVDYITSSGLPYGMSVPPKAEAEACTPGGCVYERKPLVASTTA 480
Db 834 PPLPLLOQSPATLP-----LALPGAP--APLPQ-----PONGIGRAAGAGLGA 877
OY 481 LSATESLT--LSTGSTAVAPAGLPTEFNKFLMKAVEPKSKADENTPPSEGSAL--AG 535
Db 878 MPMAEGLLGLAGSG-----GLPL-----NGLLGLNGAANPAPAS 913
OY 536 VADSGSATRQOLSKLVTSPLSMALLTNHLKSTGSPFPYVLEPLGASPSSTKLO---Q 591
Db 914 LSGAGGAPTLQLPCLNSLSTE---QQRHLLQOQEOQLQO-LQQLLASPOLTPHQVTVYQ 969
OY 592 LVEKIDROGAVAVASTASGA--PTTSAPAPSS---SASGPNCVCICLRVLSCPRLRLH 645
Db 970 MIOQIOQKRELQRLQMAAGSOLPMASLLAGSSTPLLSAGTPG---LLPTASAPPL--- 1022
OY 646 YGQHGGERPFCKVCGRAFSTRGNLRAHFVGHKTSPPARAONSCPICQKKFTNAVTLQOH 705
Db 1023 -----PAGALVAPSLGNTSLMAAAAAAANAAGGPVLTQNTN 1062
OY 706 VRMHLGGQI-----PNGSALSEGGAOAE 730
Db 1063 PFLSLGAEAGSGGPKGCTA-DKGSANQE 1091

RESULT 37
PCT-US94-04496-55
; Sequence 55, Application PC/TUS9404496
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Cannaal, Eli
; TITLE OF INVENTION: diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSER: Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca Esq., Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1242
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1093 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
PCT-US94-04496-55

Query Match 4.0%; Score 211.5; DB 5; Length 1093;
Best Local Similarity 20.6%; Pred. No. 4,1e-06;
Matches 192; Conservative 88; Mismatches 321; Indels 329; Gaps 41;

OY 3 QETGSSSRRLGCPGPAPRGDASEHNHPQVACCAQAFEDPTEFLAHONSCCTDPYMW 62
Db 289 QETSESSR-----ESKQKSSSHSLSHKCKTLSSGCVSFTTASSSSSSSSS- 338
OY 63 IIGQENPNSNSASAPRPEGHSR-SQVMDTEHSNPPDSSGSPDPPT----- 109
Db 339 --GGPQPAVSSLQSSPDSAPFKLEQPEDEYKSTPAAPASAPSPADEPPKADLFEQ 396
OY 110 -----WGPERR---GESSSQFLVAATG-----TAAGGGGLILASIKLGATPLPE 153
Db 397 KYVFSGFGPIKRFSTTTSSSGRARPSPEDYKSPHYTGSASAGTKRMALSATPVAD 456
OY 154 STP-----APPP-----PPPPPPG----- 169
Db 457 ETPETGLKEKKHKASKRSRHHGPRKGSRNKRGTTGGPAPASLPSAQLAGTATAPSPSG 516
OY 170 ---VSGS-----HLNPLILEELRVLQQRQI-HOMQTEQICROYLT 207
Db 517 GSLVSSGLGLSRTFGPSGLPSLSLESPILGAGIYTSNKDPIISHGGMIRAVCSPLS 576
OY 208 LGSIGQTVGAPASP-----SELPGTAASSTKPLPLFSPIKRAQTKTASSSSS 259
Db 577 SSLGPP-GTSALPRLSKSPPTSLPSSASISTIQVESL-----AGSTFSL 622
OY 260 SSSGAEPPKQAFHLY-HPLGSQHPFVSQGVGRSHK-----TPAP 299
Db 623 PST-----HIFGPMGAVNPDLSQAESSHPEDEDCSFRORGTSPQSLSSMSP 672
OY 300 SPALPGSTDOLASP-----HLAPFGTT-----GLLAQCLGAARGLA- 338
Db 673 ISSLPALFDQTASAPCGGQOLDPAPAGTTNMEOLLEKQDGEAGVNIWEKALHLOKE 732
OY 339 -----AASPLKPKNG-----SGELGYEVIS 361
Db 733 NORLQEQILSTAKKERIQLINVLVSFPPALPALPANGPVPGRYGLPPQAGSSDLS 792
OY 362 SLEKPGGRHKCRCAKVFSGDSALQIHLRSHTGERPYK-CNVCGNRFTTNGNLKVHFRH 420
Db 793 TSKSPGKSS-----LGLDNL-----STSEDPHSGC-----PSRSSSSLSFHS 833
OY 421 REKYPHVOMNHPVREHLVDYITSSGLPYGMSVPPKAEAEACTPGGCVYERKPLVASTTA 480
Db 834 PPLPLLOQSPATLP-----LALPGAP--APLPQ-----PONGIGRAAGAGLGA 877
OY 481 LSATESLT--LSTGSTAVAPAGLPTEFNKFLMKAVEPKSKADENTPPSEGSAL--AG 535
Db 878 MPMAEGLLGLAGSG-----GLPL-----NGLLGLNGAANPAPAS 913
OY 536 VADSGSATRQOLSKLVTSPLSMALLTNHLKSTGSPFPYVLEPLGASPSSTKLO---Q 591
Db 914 LSGAGGAPTLQLPCLNSLSTE---QQRHLLQOQEOQLQO-LQQLLASPOLTPHQVTVYQ 969
OY 592 LVEKIDROGAVAVASTASGA--PTTSAPAPSS---SASGPNCVCICLRVLSCPRLRLH 645
Db 970 MIOQIOQKRELQRLQMAAGSOLPMASLLAGSSTPLLSAGTPG---LLPTASAPPL--- 1022
OY 646 YGQHGGERPFCKVCGRAFSTRGNLRAHFVGHKTSPPARAONSCPICQKKFTNAVTLQOH 705
Db 1023 -----PAGALVAPSLGNTSLMAAAAAAANAAGGPVLTQNTN 1062
OY 706 VRMHLGGQI-----PNGSALSEGGAOAE 730
Db 1063 PFLSLGAEAGSGGPKGCTA-DKGSANQE 1091

```

RESULT 38
 US-07-945-283-2
 : Sequence 2, Application US/07945283
 : Patent No. 5352596
 : GENERAL INFORMATION:
 : APPLICANT: Cheung, Andrew K.
 : APPLICANT: Mesley, Ronald D.
 : TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
 : TITLE OF INVENTION: Involving The EP0 and LIT Genes
 : NUMBER OF SEQUENCES: 7
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Curtis P. Ribando
 : STREET: 1815 No. 5352596th University Street
 : CITY: Peoria
 : STATE: IL
 : COUNTRY: USA
 : ZIP: 61604
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/07/945,283
 : FILING DATE: 19920911
 : CLASSIFICATION: 424
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Ribando, Curtis P
 : REGISTRATION NUMBER: 27976
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 309-685-4011 ext.513
 : TELEFAX: 309-685-4128
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1958 amino acids
 : TYPE: AMINO ACID
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-07-945-283-2

Query Match 4.0%; Score 211.5; DB 1; Length 1958;
 Best Local Similarity 20.2%; Pred. No. 8.7e-06;
 Matches 196; Conservative 75; Mismatches 362; Indels 339; Gaps 43;
 QY 5 TGSSRLGSGCGRAEGRGDASEENH-----PQVCAKCAQPSDP-----TEFLAH 50
 Db 311 SGFTSHQDRPRGPTP---STSHNNHNNQRPRTSPRSTSSHQDRPGGGRPSAETNNHH 367
 QY 51 QNSCSDPRVWVITIG-----QENPSNASSASAPRECHSRQVMDTEHSNPD 99
 Db 368 Q-----DPR-----GGRRPTSSHHNNHDDPRGGGRRSPRRRSTSSSS-----SHGGRPS 412
 QY 100 SGSSGP--PRDTWGR-----BRGESSGGRVLAAT-----CTAA 132
 Db 413 TRRRPRORRRRRRRRRRPRQKISRTAGSEMTAQTLSHSEKLFHSHPMGEGBGDRGTAG 472
 QY 133 GGGGGLILSPKLGATRLPRESTRAPRRRRP--PRPGVSGHLNPLLELRLVQORQ 191
 Db 473 GEGDRDPRPR--SPRRRRRLRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 517
 QY 192 IHOMQTEQICROVULLGSLQGTGARGASPSLELPGTA-----ASTKPLRLFSR 242
 Db 518 -----GGGGGP--PRGGRRRRGKRRRACTEAAADAEE 550
 QY 243 IKPAQTGKTASSSSSSSSSGAERPKQAFHLNPLGSONHPVSGVGSRSHKPTTAP--SP 301
 Db 551 EEDGDEDEDEDRADEGREGGEGRRGA-----GGGAGESESESESSRAE 595
 QY 302 ALPGSTDDQLASPHLARPGTTGLLAOCLGARGLEAASPLKPKNGSGELGTEYVIS 361
 Db 596 GABRSASEQ-----QVGAAGVGLLVLRDGLHLDGERAAGAVAAAE--ADDLHRGRVLR 648

QY 362 SLEKPGR-----HKRCRAKVFSGDSALQILHRSHTGEPYKCNVCGNRFTTRGNKV 415
 Db 649 VLAGPPGARGPYGLHG-----AAGGADAGLEGKKVPAHGRGARGRGANGQORGLGV 704
 QY 416 HFHHRH-----EKYPVQNNPH----- 432
 Db 705 GLOQRGAGEGALGROALGAELTGEPARAAGDEDPORGAEPPAVGAVREGARVKRV 764
 QY 433 PVPEH---LDVYITS--SGLPYGMSPPEKAEAEAGTPGGVEKRLPVASTALATESLT 488
 Db 765 PEPEGALAGHVAHVHGGHDLVAGARRORDGPPGAGAH---VAHVYLAEDAQRIL- 820
 QY 489 LLSTGTAVVAPGLPTFNKFLMKAVEPKSKADENTPPGSEGSIAIGVADSGSATRQLS 548
 Db 821 -----GPGV-----QAGE-----GGLHGEAGRAHDGAVVQDGR 850
 QY 549 KLVTSLEPSMALLTNNLKSTGSP-----PPRYVLEPL-----GA 581
 Db 851 ELAALGPAHAGALGGRVQADVDVVPVHGRAVRGPLYLDGVQHDEPARRAEPRAEVLDA 910
 QY 582 SPSETSKLOLVEKIDROGAVAVASTAGAPTTSAAPSSSSASGPNOCVLCRLVLSGPR 641
 Db 911 GEAEVPRREQ-----QHPLGEADVGAP---GPVP-----GPG-----VRV 944
 QY 642 LRLHVGNGER---PFCKKVCGRAFSTRGNLPAH--FVGHKTSPAARAQNSCPTQCKFT 697
 Db 945 RRAEAVGEGGEGQREARAAARVPGRGALGLGAEELLVGQR----- 985
 QY 698 NAVTLQOHVMIHG--GQIPNGSALSSEGGAAQENSSEQSTASGPGSFPPOQOOPSPE 756
 Db 986 ---VVEHHNANHVLGVLPHRPGGAABERG-----ANGPARCGDVRGGRRVGE 1030
 QY 757 EMSDEEEDDEEEDVTDSDSLAGRSGESGGEKAISYRGSESEVSCAEDEEYATVAAPTT 816
 Db 1031 RRAPFGEEDLVHNEGAGHLGRAVGGEGRRGGRPRVGLAGR---DAAEAAGRGVLG--- 1083
 QY 817 VKEMDSNEKAPQ 828
 Db 1084 ---HGPERAPE 1091

RESULT 39
 US-09-121-321-16
 : Sequence 16, Application US/09121321
 : Patent No. 6090783
 : GENERAL INFORMATION:
 : APPLICANT: Saiga, Akihiko
 : APPLICANT: Orita, Satoshi
 : APPLICANT: Igarashi, Hisanaga
 : APPLICANT: Okumura, Kouichi
 : APPLICANT: Sakaguchi, Gaku
 : TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF
 : TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN
 : NUMBER OF SEQUENCES: 20
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: FISH & NEAVE
 : STREET: 1251 Avenue of the Americas
 : CITY: New York
 : STATE: New York
 : COUNTRY: USA
 : ZIP: 10020
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/121,321
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/933,803

FILING DATE: 19-SEP-1997
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: SHGN-12CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 671 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
S-09-121-321-16

Query Match 3.9%; Score 208.5; DB 3; Length 671;
Best Local Similarity 17.6%; Pred. No. 3.4e-06;
Matches 135; Conservative 46; Mismatches 186; Indels 399; Gaps 24;

17 EPAERGGDASEE--HHPOVCARCAQAFSD--PTEFLAHONSCCTDPVAVYIIGQENPS 71
b 280 EPLVPADASSQVKREDTLCVGRGLERAIPTESI-----TDSPI-----S 322
Y 72 NSSASSAPRPEGH-----SRQVMDTEHSNPPDSGS-----SGPDDPTWGPGR 114
b 323 AODLSRIKQEHQCVMDQDLADRIDPTDPNSESLSIAHDLSWIKOEQYP--WGPR- 380
Y 115 RGEBSGGFLVAATGTAAGGGGGLILASPKLGATPLPESTPADPPPPPPPGVGSQH 174
b 381 --DSMDGEL-----GLDSCPDSLLMWK-----NPPAPQPQP----- 412
Y 175 LNIPLILELRLVLOQRQIHOMQMTBOICRQVLLGSLGQTVGAPASSELGTGAASSTK 234
b 413 -----OROPPOLOSOPOPSL-----PPIVAENPG--GPPSR 445
Y 235 PLPLPFSPIKPAQTGKTAASSSSSSSGAEPFKQAFHLYHPGLSGHFFSVGVGRSHK 294
b 446 GLLDGDFVLPDGER-----SGEAPP-----GDDR--STGGGG----- 476
Y 295 PTPAPSPALPGSTDLIASPHLAFPTTGTGLAOCUGAANGLEAAASBGLKPKNGSEL 354
b 477 -----GDGG-----GGGGAEGGTAG-----GGCGSC 499
Y 355 GYGEVVISLEKPGGRHK---CRFCAKVFSGDSALQIHLRSHTGERPKVCNCGNFTTRG 411
b 500 CPGGLRSLRLHGAARSKYSCDEGKSFVGRKSLIIHRSHTKERPYECACAEKSFNCHS 559
Y 412 NLKVFHRRHREKYPVQNPAPVPEHLHYITSSGLPYGMSVPRPEKAEEAGTPGGVER 471
b 560 GLIRHOMTHRGERP-----GDR----- 574
Y 472 KPLVASTALATESLTLTSTGTAVAPGLPTFNKFLVAKAVEPKSKADENTPPGSEGS 531
b 575 ----- 574
Y 532 AIGAGVADSGATRMQLSLVLSLPSMALLTNHLKSTGSPFPVYLEPLGASPSSETSKLQO 591
b 575 ----- 574
Y 592 LVEKIDROGAVAVASTAGATTTAPAPSSASGPNOCVICTRLVSCPRALRLHYGOHGG 651
b 575 -----KSCSECKTYSKREHLQNHORLHTG 598
Y 652 ERPPKCYCGRAFTGRNMLRAHFVGHKTSAPARAONSCPTICOKKFTNAVTLQOHVRLHIG 711
b 599 ERFPQCALCGKSFIRKQNL---LKHQRIHTGERPYTCGCGKSFYRKESILKDHRLRVHSG 654
Y 712 GQIPNGSALSSEGGAAENSSSQSTASGSPFPQSQSPSPREE 757
b 655 G-----PpPpGA---PpPpLPPPPPERD 671

RESULT 40
US-08-933-803A-16
Sequence 16, Application US/08933803A
Patent No. 6218522
GENERAL INFORMATION:
APPLICANT: Saiga, Akihiko
APPLICANT: Orita, Satoshi
APPLICANT: Igarashi, Hisanaga
APPLICANT: Okumura, Kouichi
APPLICANT: Sakaguchi, Gaku
TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF
TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
City: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,803A
FILING DATE: 19-SEP-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: SHGN-12CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 671 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-933-803A-16
Query Match 3.9%; Score 208.5; DB 4; Length 671;
Best Local Similarity 17.6%; Pred. No. 3.4e-06;
Matches 135; Conservative 46; Mismatches 186; Indels 399; Gaps 24;
17 EPAERGGDASEE--HHPOVCARCAQAFSD--PTEFLAHONSCCTDPVAVYIIGQENPS 71
b 280 EPLVPADASSQVKREDTLCVGRGLERAIPTESI-----TDSPI-----S 322
Y 72 NSSASSAPRPEGH-----SRQVMDTEHSNPPDSGS-----SGPDDPTWGPGR 114
b 323 AODLSRIKQEHQCVMDQDLADRIDPTDPNSESLSIAHDLSWIKOEQYP--WGPR- 380
Y 115 RGEBSGGFLVAATGTAAGGGGGLILASPKLGATPLPESTPADPPPPPPPGVGSQH 174
b 381 --DSMDGEL-----GLDSCPDSLLMWK-----NPPAPQPQP----- 412
Y 175 LNIPLILELRLVLOQRQIHOMQMTBOICRQVLLGSLGQTVGAPASSELGTGAASSTK 234
b 413 -----OROPPOLOSOPOPSL-----PPIVAENPG--GPPSR 445
Y 235 PLPLPFSPIKPAQTGKTAASSSSSSSGAEPFKQAFHLYHPGLSGHFFSVGVGRSHK 294
b 446 GLLDGDFVLPDGER-----SGEAPP-----GDDR--STGGGG----- 476
Y 295 PTPAPSPALPGSTDLIASPHLAFPTTGTGLAOCUGAANGLEAAASBGLKPKNGSEL 354
b 477 -----GDGG-----GGGGAEGGTAG-----GGCGSC 499
Y 355 GYGEVVISLEKPGGRHK---CRFCAKVFSGDSALQIHLRSHTGERPKVCNCGNFTTRG 411

Db 500 CPGGLRSLHLHGRRSKPYSCPECGSKSGVRRKSLIIHRSHTKERPYECACECEKSFNCBS 559
QY 412 NLKVHHRHREKYPHVOMNPPHPEHLDYVITSSGLPYGMSVPEKAEDEAGTPGGVER 471
Db 560 GLIRHOMTHRGERY----- 574
QY 472 KPLVASTTALSTELTLLSTGSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSEGS 531
Db 575 ----- 574
QY 532 AINGVADSGSATRMQLSKLVTLSPSMALLTNHLKSTGSPFPYVLEPLGASPSSETSKLOQ 591
Db 575 ----- 574
QY 592 LVEKIDRGAVAVASTAGAPTSPAPSSSASGPNOCVLCRLVLSCPRALRLHYGOHGG 651
Db 575 -----KCSECEKYSRKELQNHORLHTG 598
QY 652 ERPPKCKVCGRASTRGNLRAHVEVGHKTSAPAARQNSCPICQKFTNAVTLQGHVRMLG 711
Db 599 ERPPOCALCGKSFIRKQNL---LKHORIHTGERPYTCGECGKSFERYKESLKDHLRVHSG 654
QY 712 GOIPNGGSALSEGGAQOENSSEQSTASGPGSFQPOSOQPSPEEE 757
Db 655 G-----PGPGA---PROLPPPERD 671

Search completed: January 13, 2003, 15:23:49
Job time : 49.9776 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein protein search, using sw model

Run on: January 13, 2003, 15:17:45 : Search time 22.4664 Seconds
(without alignments)
4287.595 Million cell updates/sec

Title: US-09-988-117-3
Perfect score: 5301
Sequence: 1 MAOETGSSRLGCGCEPAE.....SSIPSGLSPPRRKDDPTMP 1002

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1282	24.2	1323	2 T30253	spalt protein - mo
2	1265	23.9	1350	2 T30341	zinc finger protei
3	1142.5	21.6	1061	2 JC7116	Xsai-3 protein - A
4	759.5	14.3	1355	2 S40022	spalt protein - fr
5	742.5	14.0	1263	2 T13805	spalt-related prot
6	712	13.4	1402	2 S42748	finger protein - f
7	377.5	7.1	744	2 T20969	hypothetical prote
8	349	6.6	2282	2 T42717	DNA-binding protei
9	330	6.2	1173	2 S70620	prockr2 - chicken
10	322.5	6.1	744	2 S70619	finger protein bow
11	317.5	6.0	727	2 B60310	transcription regu
12	316	6.0	1615	2 JC6510	ras-responsive ele
13	311.5	5.9	1920	2 T13893	gene handsight pro
14	307.5	5.8	1891	2 T13594	hypothetical prote
15	306.5	5.8	654	2 A57785	finger protein ZNF
16	300.5	5.7	644	2 T46277	hypothetical prote
17	300.5	5.7	2232	2 T34434	hypothetical prote
18	292	5.5	624	2 S41688	DNA-binding protei
19	292	5.5	1184	2 G01763	finger protein - hu
20	290	5.5	651	2 B32891	finger protein 2,
21	290	5.5	1350	2 S00647	finger protein - A
22	287.5	5.4	982	2 T43676	hunchback-related
23	286.5	5.4	1173	2 T31421	C-terminal domain
24	283.5	5.3	2688	2 S50832	alpha-A-crystallin
25	281.5	5.3	1184	2 S50832	actophin-1 - human
26	281	5.3	1191	2 S35305	zinc finger protei
27	279.5	5.3	671	2 JE0288	Kruppel-type zinc
28	278	5.2	3942	2 T42730	Bassoon protein -
29	277	5.2	2715	2 T13049	eyeld - fruit fly

30	271	5.1	1060	2 S33641	homeotic protein z
31	269	5.1	1042	2 A31591	transcription regu
32	266.5	5.0	615	2 S06546	finger protein (cl
33	265	5.0	803	2 S26823	zinc finger protei
34	263.5	5.0	686	2 A34612	zinc finger protei
35	263	5.0	856	2 A53503	B-lymphocyte-induc
36	262.5	5.0	710	2 I48668	zinc finger protei
37	262	4.9	604	2 S05447	finger protein gla
38	261.5	4.9	536	2 S06548	finger protein (cl
39	261.5	4.9	675	2 S51037	zinc-finger protei
40	261.5	4.9	2500	1 WMHUDE2	HIV-EP2 enhancer-b
41	261	4.9	428	2 S03677	finger protein (cl
42	261	4.9	581	2 A49073	HSN motor neurons
43	259.5	4.9	789	2 A39564	transcription repr
44	259.5	4.9	794	2 S59069	Z13 protein - mous
45	257.5	4.9	594	2 JC5146	arylphorin gene-sp
46	257.5	4.9	3938	2 T42761	Bassoon protein -
47	257	4.8	2529	2 A56923	transcription fact
48	256.5	4.8	485	2 A40751	finger protein MZF
49	255	4.8	553	2 S22954	finger protein zfp
50	255	4.8	908	2 T16057	hypothetical prote
51	254.5	4.8	395	2 G02075	transcription repr
52	254	4.8	439	2 S06556	finger protein (cl
53	254	4.8	591	2 S65088	finger protein XPO
54	253	4.8	494	2 A42170	zinc finger protei
55	253	4.8	543	2 E88280	protein egl-43 lim
56	250	4.7	411	2 S10245	finger protein, te
57	250	4.7	2578	2 A56922	transcription fact
58	249.5	4.7	1706	2 I84499	zinc finger protei
59	248	4.7	595	2 JC7779	Kruppel-associate
60	246.5	4.7	1051	2 A60191	oncogene Evi-1 - h
61	246.5	4.6	427	2 T49603	transcription regu
62	245.5	4.6	459	2 I38600	zinc finger protei
63	245.5	4.6	477	2 A47236	zinc-finger protei
64	245	4.6	1042	2 S41705	Evi1 protein - hum
65	244	4.6	542	2 A54661	zinc finger protei
66	244	4.6	614	2 JH0500	zinc finger protei
67	244	4.6	2717	2 A34203	DNA-binding protei
68	243.5	4.6	487	2 JC5076	myc-associated zin
69	243	4.6	582	2 S08686	finger protein ZFP
70	242	4.6	488	2 S47072	finger protein HZF
71	241.5	4.6	399	2 S47071	finger protein HZF
72	241.5	4.6	3190	2 T13828	CREB-binding prote
73	241	4.5	693	2 I37570	zinc finger protei
74	240	4.5	457	2 C57785	zinc finger protei
75	240	4.5	701	2 T14757	hypothetical prote
76	239	4.5	509	2 JH0501	zinc finger protei
77	239	4.5	636	2 I48689	gene NK10 protein
78	238.5	4.5	456	2 T08674	probable finger pr
79	238.5	4.5	3869	2 A48205	ALL-1 protein +GTE
80	237.5	4.5	625	2 D87793	protein C27A12.2
81	236.5	4.5	480	2 JC7812	BCL6 homologous z1
82	236.5	4.5	688	2 A56360	zinc finger protei
83	236.5	4.5	2783	1 A41948	alpha-fetoprotein
84	236	4.5	1721	2 I38902	retinoblastoma bin
85	234	4.4	589	2 I38598	zinc finger protei
86	233.5	4.4	540	2 B57785	zinc finger protei
87	233.5	4.4	546	2 I49636	DNA-binding protei
88	233.5	4.4	634	2 T47156	hypothetical prote
89	232.5	4.4	420	2 S06579	finger protein (cl
90	232	4.4	1117	2 JC4934	delta-crystallin/E
91	231.5	4.4	1634	2 T26517	hypothetical prote
92	230.5	4.3	753	2 S48059	metal-regulatory t
93	230.5	4.3	1032	2 T34433	hypothetical prote
94	229	4.3	572	2 I39311	Kruppel-type zinc
95	229	4.3	899	2 B48586	suppressor of hair
96	229	4.3	1151	2 T18535	high molecular mas
97	228.5	4.3	728	2 A54603	transcription fact
98	228.5	4.3	1585	2 T31611	hypothetical prote
99	228	4.3	577	2 S72227	finger protein sob
100	228	4.3	3968	2 A44265	triflorax homolog

ALIGNMENTS

RESULT 1
330253
msalpt protein - mouse (fragment)
A:Alternate names: zinc finger protein msal
S:Species: Mus musculus (house mouse)
D:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
Accession: T30253
Kott, T.; Kaestner, K.H.; Monaghan, A.P.; Schutz, G.
Mech. Dev. 56, 117-128, 1996
Title: The mouse homolog of the region specific homeotic gene spalt of Drosophila is
A:Reference number: Z20791; MUID:96391179; PMID:8798152
A:Accession: T30253
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1323 <OT>
A:Cross-references: EMBL:X97581; NID:g1296844; PIDN:CAA66196.1; PID:g1296845
A:Gene: msal
A:Function:
A:Description: may play an important role in the development of the nervous system
Query Match 24.2%; Score 1282; DB 2; Length 1323;
Best Local Similarity 31.3%; Pred. No. 2.1e-53;
Matches 374; Conservative 133; Mismatches 329; Indels 360; Gaps 45;
4 ETGSSSRLLGPGCGPAERGGDASEHHPOVCAKCAOFSPTFEFLAHONSCCTDPPVMVI 63
:::|||||
9 DSGSESRSG-----SEE--TSVCEKCAEFKMADELQHKKTCTKNPLVLIV 53
:::|||||
64 IG-----GOENPSSNSAS--APREGHSRSQVMDENH-----NPPDSSSGP 105
:::|||||
54 HDEPAPSPEDFPFPPSPASSSDRTESEVAEAPREGSEVKAATKASMEVSTDKGP 113
|
106 PPPTWGPERRGEESSGQFLVAATGTAAGGGGLILASPKLGATPLPESTPAPPPPP 165
|
114 PGP-----SVPPPPALPPQ 128
|
166 PPP-----GYS--GHNLPLLELRLVQ 188
|
129 PEPAAFSMPSTNWTLETLSTKVAVAOFSQAGAGGTGAGSGVAVALPMILEQALQ 188
|
189 QRQIHOMQMEQICROYLLLSLGGTVGAPASSEL-----P 225
|:|||||
189 QQQIHQLLEQLRHSQV---GLMSQPGPPLKPSASAPGNTSVOLGLTPHAAQLLSGP 245
|
226 GTGAASSTKPLPLFSPIKPAQTKTFTASSSSSSSGAPPPQAFHLYHPLGSQHP-- 283
|:|:|:|
246 ATASAGSGSTLPAFD--GPQHLISQASGTSTPCSTSAAPRDSG---AHPACSTGPAP 299
|
284 ----FVVGCGRSKRP--TP-----APSPALGOSTQOLIASPHLAFPGT 321
|:|:|:|
300 GAVAAASSTVGNAAVOPONASTPPALGPGPLLSSASNLPNLLPQT-----SSSVTFPNP 354
|
322 TGLLAQCLGAARGLLEAAASGLLKPKNSS--GELGXEVISSLEKPGCRHKCRCAKVG 380
|:|:|:|
355 LYSIAA---TANALDPLSA--LMKHKKKPPNVSVPFERKASADPFKHKCRCAKAVFG 408
|
381 SPSALQIHLSRHTGERPKYKCVNCGNRTTGNLKVHFRHREKYRHVOMNHPVPEHLDY 440
|:|:|:|
409 SPSALQIHLSRHTGERPKYKCVNCGNRTTGNLKVHFRHREKYRHVOMNHPVPEHLDY 468
|
441 VITSSGLPYMSVPERK-----AEEAGTP-----GGVEKPLVASTALASLTLIS 491
|:|:|:|
469 CPTCSGIPYGMSLPERKPVTTMLDSKPVLPVPTSVGLQLPPTVPGHNHYDSSITPVS 528
|
492 -----TGTSTAAVAPGLPTFNKFLMKAVEPK-----SKADENPPGSE---GS 531
|:|:|:|
529 RSPQRPSPASSECTSLSPGLNNTESGITVRPSPQPLGLGSPSLTKAPVSLPCTSTRTGD 588
|:|:|:|
532 A-IAGVADSG---SATRMQSLKLVTSL--PSWALLTNHLKSTGSFPFVYLEPLGASPS 585

Db 589 APYVGQVSGSLPTSAANAVIDSACTSLSGPLPAVSDQFKA--QFPFGILDSM--QTSE 644
|:|:|:|
Qy 586 TSKLOQLVEKIDROGAAVAVASTAGAPTTASAPSSASAGPNOCVLTLYISCPRALFLH 645
|:|:|:|
Db 645 TSKLOQLVENIDKK-----MTDPNOCVICHRYLISCSALMKH 681
|
Qy 646 YGQNGERPRPKCKVCGRAFSTRGNLRAHVPGHKTPSPAARONSCPTCKKFTNAVLQOH 705
|:|:|:|
Db 682 YRTHGERPRPKCKIGRAFTTKGNLKTNGHGRKPLRVOHSCPTCKKFTNAVLQOH 741
|
Qy 706 VRNHLGQIPN-----GGSALSEGGAQENSSSQ-----ST 737
|:|:|:|
Db 742 IRNHHMGQIPNTPLPEGLQEMADLPPDEKNAFTLSSDDLDIDENMEDESELKDTASD 801
|
Qy 738 ASGP-----GSFPQPSQSPSEEMSE-----EEDEDEEE 769
|:|:|:|
Db 802 SSKPLTYSGSCSP-----PSPPSVSIISIALENQMKMIDSVANCOQLAMLKSVENGSGE 855
|
Qy 770 EYVTDEDSLARG--SESGEKAISVAGDSEEVSAEEVATSVAPPTYKENDSNKA 826
|:|:|:|
Db 856 SDRLSNDSSSAVGDLESRSAGSPALSESSSQALSPAHNGESFRSKSPGLGHQEDPQEI 915
|
Qy 827 PQHT--LPPPPPPNDNDHPQPMEOGTSDVSGAMEBEAKLEGISSPMALTOEGEGSTP 884
|:|:|:|
Db 916 PLKTERLDSPPPGPN-----GGALDITAGHP-----RP 945
|
Qy 885 LYBELNLPEA-----MKDQESSGRKACEYVGOSFPQTALAEHOKTHPKDGLPTCV 938
|:|:|:|
Db 946 LIKE-----EAPFSLPLFSRERKQAS--TVCGVCGKPFACKSALEIHYRSHTKR--RFVCT 999
|
Qy 939 FCRQGFLDRTATLKKHMLAHNVOPFPAHPQONATLSLVPGSSSTIPSGLSPPF 994
|:|:|:|
Db 1000 VCRGCGSTMGNLKOHLLT--HKLKL-----PSQVEDPNTLGPSSHSTPSSAPAP 1049
|
RESULT 2
T30341
zinc finger protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
A:Accession: T30341
R:Holleman, T.; Schuh, R.; Pieler, T.; Stick, R.
Mech. Dev. 55, 19-32, 1996
A:Title: Xenopus xsal-1, a vertebrate homolog of the region specific homeotic gene sp
A:Reference number: Z20832; MUID:96317243; PMID:8734496
A:Accession: T30341
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1350 <HOL>
A:Cross-references: EMBL:L46583; NID:g1235930; PID:g1235931; PIDN:AAC4223.1
C:Genetics:
A>Note: Xsal-1
Query Match 23.9%; Score 1265; DB 2; Length 1350;
Best Local Similarity 30.0%; Pred. No. 1.4e-52;
Matches 372; Conservative 155; Mismatches 335; Indels 378; Gaps 44;
4 ETGSSSRLLGPGCGPAERGGDASEHHPOVCAKCAOFSPTFEFLAHONSCCTDPPVMVI 63
:::|||||
50 DSGNBSRSG-----SEE--TNVCEKCAEFKMADELQHKKTCTKNPLVLIV 94
|
Qy 64 IGG-----OENPSSNSASAPREGHSRSQVMD----- 91
|:|:|:|
Db 95 NDDVAAPVEEVEPSPASS--PSNHAESSETAEENIOVENNDTCIDKTEKEEPEMEVE 151
|
Qy 92 -TEHNSNPDSGSGPDPPTWGPERRGEESSGQFLVAATGTAAGGGGLILASPK----- 144
|:|:|:|
Db 152 ITEKNYRPOEASDPAATPL--POLRPPSSMTNYNPNNTVTL-----TLQSTKVAVAQF 204
|
Qy 145 -----LGATPLPESTPADPPPPPPPPPGVSGHLNPLILEELRLVLOQROIHQOMT 198
|:|:|:|
Db 205 SQNAOCVGSTNATAATA-----MAIPMLIEQLMALQOOQIHOQLI 246

[illegible]

RESULT 4
40022

palt protein - fruit fly (*Drosophila melanogaster*)
 ;Species: *Drosophila melanogaster*
 ;Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000

;Accession: S40022
 ;Kuhnlein, R. P.; Frommer, G.; Friedrich, M.; Gonzalez-Galtan, M.; Weber, MBO J. 13, 168-179, 1994
 ;Title: spalt encodes an evolutionarily conserved zinc finger protein of n
 ;Reference number: S40022 | PMID:94139659; PMID:7905822
 ;Accession: S40022
 ;Molecule type: DNA

```

;Residues: 1-1355 <KOE>
;Cross-references: EMBL:X75541; NID:9414106; PDB:3CA55229,1; PDB:92568394
;Note: the authors translated the codon GAC for residue 51 as Ala
;Note: mRNA sequencing has also been done
;Genetics:
;Gene: sal; spalt
;Cross-references: FlyBase:FBgn0004579
;Introns: 51/1; 1329/3 1355/2
;Keywords: zinc finger

```

Query Match	14.38;	Score 759.5;	DB 2;	Length 1355;
Best Local Similarity	22.78;	Pred. No. 1.2e-28;		
Matches 312; Conservative	141;	Mismatches 415;	Indels 505;	Gaps 44;

```

xy 3 QETGSSRLGGPCGEPRAREGRGDASEHHNPQYCAKCAQAFSDPPEFLAHQNSCCCTDPPEVM 62
xy 50 KDIGSOOEENGSGSPLTATTATTAPSRNPE-----PEEEQPEOSTSEOSI---PEOST 100
xy 63 IIGGOENPNSMSASAPRPEGHSRQVMDTEHSN---PPDSGSGGPPDPWGPERRGEES 119
xy 101 PDHQLENDIKSEAKSELIEPVEDNNRRVMTKPSSEBREPFASSGMPSSPY--AFASAEAA 158
xy 120 SGQFLVAATGTAAAGGCGGLILASPKLGATP-----LPESRPAAPPPEPPP 166
xy 159 ATE-----RTPEKEKENVDEVDMPEAPPSAVSTENTYL 194
xy 167 PPGVSGGHLNIPILLELR-----VLQ 188
xy 195 PGAGAG-----PYTLEAIQNQMMAIAGRAAKTIANGSNGADNEAMKQLAFLQOTLEMLQ 249
xy 189 OROIHOMQMTEOICROYLL-----LGSIGQTVGA 217
xy 250 OQQLFQIQLHQLOSQALINAQOKOEDEPTEADQEQDQEQETPTVEEERLIADMLROKA 309
xy 218 PASPSLPGTGAASSTPELPLPFSRIKAQOTKTATASSSSSSSSGAEPPKQAFHLYHP 277
xy 310 FARMAE-----AKAROHLLINAGVPLR-----ESSGSPAESLKRRREHDI-- 348
xy 278 LGSQHFPVSGGVGSHKPTPA-----PSPALPGSTD--QLIASPH--LAFPGTG 323
xy 349 --ESQPRKRTSLDNTFKADVTADQALAKLEKEMENTPLPFSSDLASSITTNHDDLPENSLD 406

```

QY	324	LLAOCGLGANGLEMAASVGLLK-----PKNGSGELGGEVLSLEKEGGRHKR	373
QY	407	LLORK---AOEVLDSASOGILLANSMADDPARGEKSGEGK-----GRNEPEFKHKR	454
QY	374	FCAGYFGSDSALOIHIRSTGTERPYKANCVGNFTTJRGMLKVFHNRHREKVRPHOMPHR	433
Db	455	YCGYVFGSDSALOIHIRSTGTERPYKANCVGNFTTJRGMLKVFHNRHREKVRPHOMPHR	514
QY	434	VPEN-----LDYVITSSGGLPYGMSVPERKEAEEBAGTRPG--GYER--KPLVASTAL	481
Db	515	IREHMDKFNRLDDQMSPTDSSPNHSRAPPRLGSAASAPPAERGLONLXRPMEILKSL	574
QY	482	SATESLTLSTGTSNAV---ABGLPTFNKFLYLMKABERKSKADENPRGSGS-----	531
Db	575	GAAPHOYFROELPYDLKRPSLOLDEDEROVUKNEPYEBKOREHEDEMAECSEBERPL	634
QY	532	-AIGVADSGSATRMOLSK-----LVPLSRPMALITNLKSTGSEPPRYVLEPL----	579
Db	635	PLEVATKEERVEDEQVOKOEDHRIERPRKRPSPSEHSRPHNNHNSHNGYRPPVQIPDRA	694
QY	580	---GASESETSKILOL-----VEKIDROGAVALASTASGAR	612
Db	695	LMHPQSPGSGOHLNHLPRPOLPRREDFEAFERPLNFTAKMLSPENHSFVRSRAGAL	754
QY	613	TTSAPAR-----	619
Db	755	PRGYVPRPHNHRNHAESPFPNPIKHEMAALLPRPHSNDNSWENFEIVSNTCETMKLKL	814
QY	620	---SSASGPNOCVILRYLVCSPRALRYHGOHGEREPFKCYKSGRASTRGNLRAHFVGH	677
Db	815	MKNKISIDPNOCYUODRYLVCSCSALOIMRTIHNGERPFKRIGORATITGKLNKLTMAVH	874
QY	678	KTSPARAANSCRIOCKKFTNAVTIIOAHVMNLG-----GQ	713
Db	875	KIRPRMRFHOCRPYCHKKYSNALVLOQHRLHFGERTDLTPEOIOAAEIRDRPPSMRGN	934
QY	714	IPNGSALSSEGGGAQOENSSEOSTAGPSGSPFR-----QSQOPRSPEEMSEEE	762
Db	935	FMNPRMAAAFHNGA-----LPGPGGPRGPRNHGANGALGESSESGQDDNMOCGE	985
QY	763	EEDF-----EEDF-----VTDEDSLAGR	781
Db	986	DYDDVSSNHLNSNLEFOGDRSRSDDDFKSLFPEOKLRIIDATGVAVTNPVRRPSSASH	1045
QY	782	GSESGGEKA---ISYRGDSEEV---SGAEEVA-----TVAAPRTYKEMDS-----	822
Db	1046	GHSVSTSPAPTSPVNASQVYIKRSSSPARSEASOGALDLTPRAAPRTSSSSRSRPLPEK	1105
QY	823	-----NEKAPRHTLPRPPR-----PUNLDH-----	843
Db	1106	PVSPRLSPRSPSGSSSHASANILTSLPRPVGIDCLPRGLONHNLQOONHLMQOANAFAA	1165
QY	844	-----POPMOQSTSDVSGAMEEPAKLEGISSPMAL-----TOEG	878
Db	1166	AAAGHHNHQMAALDQHOHQDLREAEABOQKAAAAAANAAAAAQAOTPRQADQOREG	1222
QY	879	-----EGTSTRIVEELNP-----EAMKK-----DPGESSGR	905
Db	1226	GRGAGRPRLMLGARPRFGMRPLPRPATTONMCMANMOIAOSVMPAARFNPALSGV	1285
QY	906	K---ACEVCGSGFPYOTALEHDKTPRKQSLRTTCYVCRGQGLDRATLKTKML	955
Db	1286	RGSITTCGICQYKFFRCHSALEIHNRSTKERN--FKCSICDGEFTTKGNLKMOML	1337

RESULT 5
T13805

C:spalt-related protein - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13805
C:Barilo, R.; Shea, M.J.; Carulli, J.; Lipkow, K.; Gaul, U.; Frommer, G.; Schuh, R.

submitted to the EMBL Data Library, August 1996
A:Description: The spalt-related gene of Drosophila melanogaster is a member of an ancient
A:Reference number: 217768
A:Accession: T13805
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1263 <BAR>
A:Cross-references: EMBL:Y07653; NID:e1004366; PID:e267584; PIDN:CAA6937.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0000287
A:introns: 862/1; 915/3; 1003/1; 1080/3; 1146/3

Query Match 14.0%; Score 742.5; DB 2; Length 1263;
Best Local Similarity 23.5%; Pred. No. 7.4e-28;
Matches 279; Conservative 143; Mismatches 379; Indels 385; Gaps 41.

QY 71 SNSASAPRREGHSRSQVMTENHPDSSGGPPDPTWGPERRGEGSGQLVLAATGT 130
DB 54 SNNNGTE-PQMEADAVE-SDTERETAERGEGOEPE-----NSNEALDLTSSGGR 104
QY 131 AAGGGGLI-----LASPKLGATPLRPSTPARPPPPPPPPGVSGHINIPIL 181
DB 105 ESILPSSGVHLEALQHTKVAQAQPAATMAGNHQA-----DLMAVQ 146
QY 182 EELRLVLOQRQIHQOMTEQICROV-----LLLG-SLGQTVGAPASPSLPGTGAASSTK 234
DB 147 STTFNVORQHLMOQLQIHLQSQLKRAEAAALGRHSHDEEEREPERPKQPTNGLK 206
QY 235 PLPLPLSPIRPAQCGKTTASSSSSSSGAEPKPAFHILYHLSGCHPSS-----VGG 288
DB 207 EELRLVLOQRQIHQOMTEQICROV-----EPEGYSMMQDISSTLASS 260
QY 289 VGRSHKPTPAPS-----PALGSTDQLASPHLAPGTGTGLLAOCGAARGLPAAASPG 343
DB 261 IITNHDPRPAPNENCLEMLORRTEVIDS-----ASQSIHAQMODE----- 303
QY 344 LKPKNGSGELGVEVTSLEKPGG--RHKRCFCAKVGSDALSQIHLKSHNGERPKC 400
DB 304 -----YSEYASKEAQNSRGEIFKHCKYCKGIFGSYALQIHLKSHNGERPKVC 351
QY 401 NVCGNRTTGNLKVHHRHHEKYPHVQMNPRPVREHLVITSSG-----LP 448
DB 352 NVCGSFTTGNLKVHQRHQTQIFPPMLLPQVAPN-----VHSGGGQVQGEQPIRLP 406
QY 449 YGASVPEPEKA-----EAGTPGGGVERRPLVASTTALATSTLTLSTSTAVA 499
DB 407 FAPPAVAVGEOQHONOVEEPEEIPVPOAEDLSKPMVKEK--KSHSPVECVKTPKEVKT 464
QY 500 PGLPTPNK--FVLMKAVEPKSKADEN-----TPPGS-----EGSALAGVADSGS 541
DB 465 ASLPSSSEKPEKEISKPVYTSRRNGSVKRRQTSVAVSQEDRERDLVEHNNIAKLVRSS 524
QY 542 ATR-----NQLSLVYLSLPSVALLTNLKSSTGFPFPVYLEPLGASPSSETSKIQL 592
DB 525 ASHESQPAEYSLAQMERTIDK--SWEDLIEIDKT-----SETSKIQL 565
QY 593 VERIDROGAVAVASTAGAPTTAPAPSSASGPNOCVTCRLVLSQPRALRLHVGQGE 652
DB 566 VQIE-----NKLIDPNOCIFQKQVMSCRSSIQMNRTHTGE 602
QY 653 RPKCYAVCGRAFSTRGNLRAHVGHTSPAPARAQNSCPIQCKKFTNAVTLQOHVRLH-- 710
DB 603 RPRRCICGCAFAATKGLKAMSIHKIKPRMRSQFCVCHQKFSNGIILQOHIRITMD 662
QY 711 ---GGQ-IPNGSALSPSGGAQENS-----EGSTASPGSPQ 746
DB 663 DSGGGGVPAANPGEAELGIEDQNSNSLGTSDTLDFSTTISDHSGQRESSEGQGFDE 722
QY 747 -----POSOQSP--EEMSEEEDEEEEDVYDENS-----LAGRSES 785
DB 723 FMTMDSTDSRDNSSAATATPHPLERERDRERKRIRINDCSDESHSPDLTGGSSES 782
QY 786 GGEKAISVRGDSB-----VSGAEE-- 806

DB 783 GEMPAMDLSPPSSNSGRIFATGLANGATGGGNGGLPMLGMPPNILLMAAREENHA 842
QY 807 -----VATSVAA----- 813
DB 843 LGHAHAKEPLLPGPLGEMGLHPRPVNCILCFKMLPLSLALESHQSEHAKEPATGHAQR 902
QY 814 -----PTTVKEMDSNEKARQHTLPPRP--PRPDNDHQR 846
DB 903 PHCDAGSPYCAKLTLPNPLFAKKRPPSSSSSGEKLPESSNPPFPAMNPATPIKEDPD 962
QY 847 MEQCTSDVSGAMEEBAKLEGISSPMAALT--OEGETSTPLVE-----ELMLP 892
DB 963 QEQ-----LMVEGASAGESGTCATSNYPQADGADBSLKMQLAHNFRPASPLDQ 1015
QY 893 EA-WKKPDGESS-----GRKACVCGSGSPPTQALAEHOKHTHPKQDPLFTVCRQGL 945
DB 1016 QALMSAGPPTSLDPPVNNKHFCVCRNRNFSSSALQIHRTHTGDKP--FQCNCQKAF 1074
QY 946 DRATLKKHMLLAHQVPPFAPHPQNIATLSLVPKSSSIPSGLS 991
DB 1075 TKGNLKVHM-GTHMTNPTSRGRRMSLDPMRGPVSGGHPGSS 1119

RESULT 6
S42748
finger protein - fruit fly (Drosophila virilis) (fragment)
C:Species: Drosophila virilis
C:Date: 20-Oct-1994 #sequence_revision 26-May-1995 #text_change 24-Sep-1998
C:Accession: S42748
R:Schub, R.
Submitted to the EMBL Data Library, November 1993
A:Reference number: S42748
A:Accession: S42748
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1402 <SCH>
A:Cross-references: EMBL:J27444; NID:g426461; PID:g426462
C:Genetics:
A:Gene: FlyBase:Flyr/salm
A:Cross-references: FlyBase:FBgn0013137
C:Keywords: DNA binding; transcription regulation

Query Match 13.4%; Score 712; DB 2; Length 1402;
Best Local Similarity 21.0%; Pred. No. 2.3e-26;
Matches 306; Conservative 145; Mismatches 365; Indels 642; Gaps 48.

QY 7 SSSRLGGPC--GEPAERGD-----ASEHHPOYCAKCAQFSDPTEFLAHONSCCTD 57
DB 60 SASAGNSPCPSRSPQOHEDEREPEQVSDQLVPEVSAQSESELGETIENNADETNADHN 119
QY 58 PPMVVTIIGQENPSSNASAPRREGHSRSQVMTENHPDSSGGPPDPTWGPERRGE 117
DB 120 -----HNNNNNNKLVMTKRPVEHEVQNNANLASMP-----NSTTP 156
QY 118 ESSGQLVAATGTAAAGGGGGLLASPKLGATPLRPSTPARPPPPPPPPGVSG----- 173
DB 157 -----ATNAVYAGA-----RAQDFGATPVTLEAIONMOMALIAQPAKTIANGASGT 202
QY 174 -----HNLPILELRLVLOQRQIHQOMTEQICROVLLG----- 209
DB 203 DNEAMQMLARLQTLFNLQOQQLFQLOLQLOQSLALNQVONDEADELEPERED 262
QY 210 -----SLGQTVAPASPSG-----LPGTGAASSTRKPLPLPSPIRPAQTG- 249
DB 263 GEDTYEEERIAMELRQKAEARMAESKARQHLINAGVYANAP--DPSHQPPHRCRLRR 321
QY 250 -KTTAASSSSSSSGA-----EPPKQAFPLHYPLGSHQFFSV--GGVGRSHKPT 297
DB 322 LKKRREDASAKSSGASAKITGEQESSQDALNKLKEMENMPLEPGADLSSITINHDDL 381
QY 298 AFS--PALPGSTDQLIASPHLAFPGTTGLLAQCLGAARGLAFAASPGLLKPKNGSGELG 355
DB 355

382 EPNSLDLQKRTQEVLD-----ASQGL-ANMADFA 414
356 YGEVIS---SLEKPGRRKRCACAVFGSDSALOILRSHTERPYKNCVGNRTTGN 412
415 FGDKSSDGKGNPEFFKRCRCYCGVFGSDSALOILRSHTERPYKNCVGSRTTGN 474
413 LKHFHNRREKYPVOMNPHRPHENDYITSSGLPYGMVPRPEKAEAGPGGVK 472
475 LKVFQHNQKFRVHPMAPIREHND---KHPRLDQSPSSPTQSPATGL--- 527
473 PIVASTALATESITLSTGTSTAVAPGLPFENK--FVLKAVEPKSKADENTP----- 525
528 PRPSTSLTQOMPKMSF---ASSPAFGLPGIYRPMELLSKATGATGSLGHPHPR 583
526 --FGSGSALAGVAD-----SG----- 540
584 QMGGL-GAALKHTHDSQDMPTDLKSSGSPSPHEEDNTAARLPVKSELMEEKTEHTM 642
541 -SAPR-----MOLSKIVT----- 552
643 EAATRESAEMERPLLEVARIKEERIDEDOMHLOEGMOKREPLTAVATPHPOCLIPTTAA 702
553 -SLP-----SWALTNNLKS--TGS----- 569
703 AKSPRSRLQCHARLSLWCSHTTSMHACAVLIGSQTHLDQLPTPDNVPRTPMQRDEFA 762
570 --FFPY-----VLEPLG----- 580
763 ERPLNLTSTKTDHSPKSPAGHANAHIRSPFPNPKHMAAFVPRHSDNSWENFTE 822
581 -ASPERSEKLOOLYEKIDROGAVAVASTAGARTTSAPRASSAGMOCVCLRYLSCP 639
823 VSTSTETMKLELK-----NKKISDPQCVCYCDRLSK 857
640 RALRLYHONGGERPFKCKVCGRASTRGNLRAHFVGHKTSPPAARONSCPIQKFTNA 699
858 SALOMHYRTHTGERFRCRCGAFATTKGNLKTMAVHKIRPRNFQCPYCHKYSNA 917
700 VTIQOQVHMHG-----GOIPNG----- 717
918 LVLDQHLRLHTEPTDLTPEOTIAAERIDRPRPSMIRGFNPFMAAARHFGAMPGGAGG 977
718 -----GALSSEG----- 725
978 PRGATGMPRGCHNGTGLSESSGDDLDNMDCGDDPDDISSEHLSNDRPATSDRSS 1037
726 -----GAQENSSEFQ--STASPGSF-----POPOSOOP----- 752
1038 DDFKSLLEOKLRIDPTGVNINSHORPHSAASNPNSIGSASAPSPAPTSPSSQRPSCS 1097
753 -----SPREEMSEEEDEDEEEDVTDE--DSLARGSSSGGEKATSV----- 793
1098 PVNSSCPVRSVS---FTSQGALDLTPRALPPRLASSSSSRSPYRQLLSVRRRLARSVS 1153
794 ----- 793
1154 SHRCVAVMVRALLSSQLRPSVIGIDLRPGLOHNLQOQHNLMOQOXANVAAAADQHNHQ 1213
794 -----RGDSEEV--SGAELEVATSVAPRTYKEMDSNEKARQHTLPR--- 833
1214 MOQHAALHNOEHNRBAQEOOKAAQEOOKAAQAAAAAARQSSPQ---PRRSG 1270
834 -----PRRPRLNDHPOR-----MEQSTSDVSGAMEEERAKLEGISSPMALTOE 877
1271 ESSVGPAPRAPRLISARPPFGMPRLPIRRPATTONMCMANQIAQSVMPAPRPFPLAS 1330
878 GEGTSTPLVEELNLPEAKMKDPEGSGKACAEVCGSGFPTQALAEHQKTHPKDGLFTC 937
1331 GVGSGT-----TGICYKTFPCHSALEIHTKSHTERP-FKC 1366
938 VFCROGFLDRATLKKHML 955
1367 NICDRGFTTKGNLKHML 1384

RESULT 7
T20969
hypothetical protein F15C11.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 15-Oct-1999
R:Accession: T20969
R:Wilkinson, J.
submitted to the EMBL Data Library, April 1996
A:Reference number: 219352
A:Accession: T20969
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-744 <MIL>
A:Cross-references: EMBL:271260; PIDN:CAA95798.1; GSPDB:GN00019; CESP:F15C11.1
A:Experimental source: clone F15C11
C:Genetics:
A:Gene: CESP:F15C11.1
A:Map position: 1
A:Introns: 26/71; 75/3; 108/3; 145/3; 251/1; 453/2; 500/2; 613/1; 649/1
Query Match 7.1%; Score 377.5; DB 2; Length 744;
Best Local Similarity 22.1%; Pred. No. 8.1e-11;
Matches 182; Conservative 92; Mismatches 280; Indels 271; Gaps 29;
OY 22 GGDASEBHPQVCAKCAQESDPTFEFLAHQNSCTDPVWVITIGGQPNPSSASSAPRP 81
DB 51 GGALPLEDRNSNLIHFVVFANPOFL---SLCAQ-----LGNSSRNVSSTA----- 96
OY 82 ECHSRQVMPTHEHNPDPGSG---SGPPDPWQPERRGEESGOFVAAATGTAAGGGGL 138
DB 97 -----TTSSCPLOSQSCQSFSSPALTHVLDANHEBOELFSDCVTTSSNGDI 146
OY 139 -----ILASPKGAPRLPRESTPAPRPVPPPPVGGSHLNPILLEELRVLOQRO 191
DB 147 REHKQCKTAS---RSTSVPSSTIPSSVCFLLSTPTPCL-QFSINESIGTSEIEDEDEE 202
OY 192 IHOMQTEQICROVLLGLSLQGTVGAPASPSLPGLGAASTKPLDLFSIKAQTKT 251
DB 203 DMDEGEHVAHQ--LFGHLQ-----GKAVPTTD-----KS 224
OY 252 TASSSSSSSGAPRPKQAFNLHLPGLSQHNPFSVGVGVRSHKPTPAPSPALPSTPOLI 311
DB 225 DDKSKMASLFNHAPRFAAFNMPPFLMRQPD-----PRADVFA 265
OY 312 ASPHLAPFTGTTGLAOCLEAGLEAASPGLLKPKNGSGELGYEIVISLEKPGGRHK 371
DB 266 AGRHNDMDDEALMEISTDEAEKIRALV-----GDKAVPTTD-----NQ 306
OY 372 CRCACAVFGSDSALOILRSHTERPYKNCVGNRTTGNLKVHFRHREKYPVOMN 431
DB 307 CILCRVLSCKSALOMHRTHTGERPFKCKICORATFTTKGNLKTMHGVHRSKH----- 359
OY 432 HVPENHDYITSSGLPYGMVPRPEKA-----EEEAGTGGGV-- 469
DB 360 -----SFRGLR--ISLRPOLAAMHQHQLAPRQRINHNPTSAASAASAAVAAQ 406
OY 470 -----ERKPLVASTALSTATE--SLT-----LSTGSTAVA--PGLPTF----- 505
DB 407 IQASOOCRCIOQRFNLNGELAVHITHEHNSLTORPRVMPRTPTRYQTFPRVPRFTTTPS 466
OY 506 -----NKFLVMAVEPKSKADENTPRESSEGSALAGVADSGATRQLSLVISLPFWA 558
DB 467 LNATDMSTQENLANILSAQLNDSB--PMTDPSV-----EERITRDPPMAISLSPS-- 517
OY 559 LILNHLKSTGSEFPVYVLEPGLASPESETSKLOOLVEK--IDROGAVA-----VAST 607
DB 518 ---NSSDSSS---VKODILSESEFEKLLKLEPRLIEQOYSTPNNKNENPLLMQ 569
OY 608 ASGAPTTSAAPSSASAGPNOCVLCIRVLSCPRALRYHGOHGERPFKCKVCGRASFSTR 667
DB 570 KMAAETEPFRPPQRPVLSKHQCGVCFKHFFSSSSSALOILHMTHTHGDKRFKCMCGRAFTTR 629

```
OY      668 GNLRAHFVGH--KTSPARA----- 685
      111: | | | | | | | | | |
Db      630 GNLKVMHGTHSMWQSPSRGRIRFDVASSVTEKPMQSPILPTSGAPGAPLMLGPNGL 689
OY      686 -----QNSCPICQKFTTNVTLQOHVNMHLGGQIPNGSSA 720
      111: | | | | | | | | | |
Db      690 SGLEMMMLMRTVCSCOKVCQSPNELQHLEHLN-----NGSSA 730

RESULT 8
742717
DNA-binding protein Rc - mouse
N:Alternate names: Ig kappa chain gene enhancer Recognition component
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T42717
R:Wu, L.C.; Liu, Y.; Strandtmann, J.; Mak, C.H.; Lee, B.; Li, Z.; Yu, C.Y.
Genomics 35, 415-424, 1996
A:Title: The mouse DNA binding protein Rc for the kappa B motif of transcription and for
a family of large transcriptional proteins.
A:Reference number: 222238; MUID:97001141; PMID:8812474
A:Accession: T42717
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2282 <MUI>
A:Cross-references: EMBL:L46815; NID:q1377885; PID:q1377886; PIDN:AA840884.1
A:Experimental source: strain BALB/c; clone T1; thymocyte, brain
C:Genetics:
A:Gene: Rc
C:Function:
A:Description: binds V(D)J recombination signal sequence and kappa B motif
C:Superfamily: HIV-Ep2 enhancer-binding protein
C:Keywords: DNA recombination; transcription factor

Query Match      6.68; Score 349; DB 2; Length 2282;
Best Local Similarity 23.18; Pred. No. 6.2e-09;
Matches 216; Conservative 100; Mismatches 316; Indels 304; Gaps 44;

OY      203 ROVLILGSLGQTVGAPAPSELPETG---AASSTKPLPL--FSPIKPAQNGKTTAASS 256
      11: | | | | | | | | | |
Db      18 RKRLLRKGALIGTVSSSAP--YRSGTTPASESATQELLATPFGSGPEKKTQ----- 69
OY      257 SSSSSSGAEPKQAFPHLYHPLGSQHDFSVGCGVGRSHKPTPA-----PSPALPGSTDQL 310
      11: | | | | | | | | | |
Db      70 --QOKPARPSEIASVHI-----SQLP-----QHPLTPAFMSPGKREHLLLEGSTWQL 114
OY      311 IASPHLAFPGTGLLA-----QCLGAARGL---EAAASPGLLKPKNGSGELGYGEV 359
      11: | | | | | | | | | |
Db      115 VDPMR---PGSGSEVAPGSHPOSQLLPASHASILPPELPGIPKVFVPRPQVSLKPAEE 171
OY      360 ISSLE---KPGGRHAKRCFCAYKFGSDALQIHLRSHTGERPYKCNVCGNRFTRGTLKY 415
      11: | | | | | | | | | |
Db      172 AHKRRKPKQR--GKICYQCSRPCAKPSVLQKHIRSHTGERPYPCGPCGCFSTKTKSNL-- 228
OY      416 HFHRRREKRPYQOMNHPVPEHLIDYITSSGLPYGMS---VPPER-----AEEE 461
      11: | | | | | | | | | |
Db      229 -----YKHKRSHAHRIKAGLASGSSSEMYPPGLEMERIPGEEFEETEGESTDSEEE 280
OY      462 AGTPGG---GVEKKPIVASTTALSATE-----SLTLLSTGSTA--VAP----- 500
      11: | | | | | | | | | |
Db      281 TGAAGGPTDVLPRKPHLILSSSLYSSGSHSSQERCSLSSTGSTPGLDPAPFAPASSE 340
OY      501 -----GLPTFNKFTVLKAVPEPKKADENT--PGSEGSALAGVADGSAATRMOLSKL 550
      11: | | | | | | | | | |
Db      341 HPLSHKPEDTHITKCLALRLSERKKLIEQOTFLSPGSKGTESGYFSSSEAEQVSP 400
OY      551 VTSLSPLSMAL-----TNHLKSTGSEFPPIVLEPLCA-----SPSETSKIQ 590
      11: | | | | | | | | | |
Db      401 NTNASTVAEIIIFGKCGRIQORTSMLASTSTOP---LLPLSEEDKPSIVLPVPRQVIE 456
OY      591 QLVEXKID--RQAGVAVASTAGAPTTSPAPSSSASGPNQC----- 629
      11: | | | | | | | | | |
```

```
Db      457 HITKLITINEAVNDSEIDSVKPRRSILTRRSVSESPKSLYRDSLSHGKTKQEOQL 516
OY      630 -----VICRLVLSCPALRLHYQONGGERPFCKVCYGRAFSTGNLRAHFVGHK 678
      11: | | | | | | | | | |
Db      517 SLQHPSSTHFVPLRLSHMPSA-----ACTISTHHNFRGSY--SFDDHV 560
OY      679 TSPARAKNSCPICQKFTTNVTLQOHVNM--HLGGQIPNGSSALSSEGGAAQDENSEQ 735
      11: | | | | | | | | | |
Db      561 ADPEVPSPKNT--PV-----FTS-----HPRMLKRNAHIELPLG-----EYSSEE 598
OY      736 STASGPGSFPOPOSOQPSPEEEMSEEEDEDEEEDVTD-----EDSLGRG 782
      11: | | | | | | | | | |
Db      599 -----PGPSSKDPF-----SKPSDEPERESULTTKTKGFKTKGANGECTTICAR 644
OY      783 -----SESGEKALISVRGDSSEVSGAEEEV-----ATSYAAPTTV 817
      11: | | | | | | | | | |
Db      645 YKRDNYEAHKRYCYSELDQITKAHSGVAHEVEKTAQAEPEFWSQMNYKLGATLETLPLRK 704
OY      818 KEMDSNEKAPQHTLPRP---PRPDNLHPQPMEOGTSVSGAMEEPAKLEGISSPMAL 874
      11: | | | | | | | | | |
Db      705 RR--KEKSLGDEEPPPAFCGPPSETAHNR-----LGSTKSPAEA- 743
OY      875 TOBEGSTPLVEE--LNLPEAMKRDGESSGRKACEVCGSPPTQALAEHOKTHPKDG 932
      11: | | | | | | | | | |
Db      744 -----SKSAPSELDPPRASSPGLPSQELQNOGRRG--EQCPKKTIVIOHTSSFESKDPPEQ 797
OY      933 PLTTCVCRGQGLDRAVLTKHMLLAHNOVPPAPHG 968
      11: | | | | | | | | | |
Db      798 P-----SGLEBDKPPAQFSSPPAPPHG 819

RESULT 9
150620
Prockr2 - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: 150620
R:Schultz, B.; Niessing, J.
Gene 148, 227-236, 1994
A:Title: Cloning and structure of a chicken zinc finger cDNA: restricted expression 1
A:Reference number: 150620; MUID:95047430; PMID:7958949
A:Accession: 150620
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1173 <SCH>
A:Cross-references: EMBL:X56805; NID:9577018; PIDN:CAA40140.1; PID:9577019
C:Genetics:
A:Gene: ckr2

Query Match      6.28; Score 330; DB 2; Length 1173;
Best Local Similarity 19.88; Pred. No. 2.4e-08;
Matches 187; Conservative 81; Mismatches 278; Indels 398; Gaps 34;

OY      1 MAOETSSSRILGPGCPGPAERKGDASEEHN-----POVCACCAQFSDPTEFLAHONSC 54
      11: | | | | | | | | | |
Db      58 LRQHTGERPYKCTSCPAKFAKSSSLRRHRHITHGERPYCASGKAFQTGSTNLROHRTH 117
OY      55 CIDPRYVAVIIGQENPENSASSAPRPRGHSRQVMDTEHSMNPDSGSGGPPDPWGPFR 114
      11: | | | | | | | | | |
Db      118 TGERPY-----ACSHCSKFTTHSSNLL-----HQ 142
OY      115 RGEESGQFLVATGTAAAGGGGLILASPKLGATPLPESTPAPRPPPPPPPGVSGH 174
      11: | | | | | | | | | |
Db      143 KHSSTRSHKRCPRKAFVDACIQKLQSHASAPL-----LPSLSPQSLSPP----- 191
OY      175 LNIPLLEELRLVQROIHQOMTE---QICROYLLLSLGQTVGAPAPSELPETGAA 230
      11: | | | | | | | | | |
Db      192 ---PLLEAVEML-----YKTECTWLSPCQEGLL--GHGRSHNQPAHTPGV--TATA 237
OY      231 SSTKPLPLPLSPKPAQTGTTASSSSSS---SSSGAERP-----KQAFHLYHPLGSG 281
      11: | | | | | | | | | |
Db      238 THRCP-----TCGRTFKNSSGLARRHSHGAEPRPYKCSOHRSFQGLAGLGHQ 286
      11: | | | | | | | | | |
```

282 HPFSVGVGRSHK-PTPAPSPALPGSTDOLIASPHLAFPGTGTLLAOCIGAARGLEAA 340
b 287 -----RGHSAETPPPPAPPTPT-----SVPS----- 308
341 SPGLLKPKNGSGELGYEIVSSLEKPGGRHKCRFCAKVFGSALOIHLSHTEGPERYKC 400
b 309 -----ERP-----YCTEGCKAFKSGSGLRKYMMDHNGERPYKC 342
y 401 NYCGRFTTGRNLKVHFRHREKYPHYQMNPHVPEHLDYVITSSGLPYGMSVPEKAEE 460
b 343 SCCKPAKFKRSSLLAIH-----QRVHTGLRAVKCP-----SCGLTFKWS----- 380
y 461 EAGTGGGVGRKPLVASTALSAATESLTLSTGTSTAVAPAGLPTFNKFLMKAVEPKSKA 520
b 381 -----SHLQYHLRLHTGERPYKCPDC-----KA 404
y 521 DENTPPGSGSAIAGVADSGSATRMQLSKLVTLSPSMALLTNHLKSTGSPFPYVLEPLG 580
b 405 EKNT-----SCLGPH----- 414
y 581 ASPSETSKLOOLVEKIDROGAVAAVASTASGAPTTAPADSSASGPNOCYICLVLSGPR 640
b 415 -----ROL-----HTGERPHACSPICGKAFYOTS 437
y 641 ALRLHYGNGGERPFPKCVCGRAFSTRGNLRAHFVGHKTPSPARAONSCPICOKKFTNAV 700
b 438 NLROHQRHTGERYACSHCKTFPHSSMLDLH-----QRTHSSAR-PRHOCPLCPRAFWAS 493
y 701 TLOOHVYRMLHGGQIPNGSGALSSEGGAOENSSSEOSTASGPGFPPOSOQSPSEEMSE 760
b 494 YLQRHLRLH-----AAGPKGSPRP----- 512
y 761 EEEDEEEEDVDYDEDSLACGSGSGEKAISVRGDSSEYSGAEEFVATSYAAPTYVEM 820
b 513 -----ALTPQRDGPVYLQAA-----LSLEYTAP----- 534
y 821 DSNE-----KAPQ-HTLPPPPPPPNL-----DHPQMEQSTSDVSGAMEEAKLEGIS 868
b 535 DAHTFLLLQTPPGQILPSPRAPOKLILPTAPQRPKQSGEPPTGQSLLVSTGTT 594
y 869 SP-MAALTOEGEGSTPLVEELNLPEAMKKDPGSSGSKRACE 909
b 595 LPTLRLOAVTAVPQGTGTGLVLQGLPEQRLHPAGIPIHGOAAVE 638
RESULT 10
570619
fingerprint protein bowel - fruit fly (Drosophila melanogaster)
Species: Drosophila melanogaster
Date: 19-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 17-Nov-2000
Accession: S70619
A: Wang, L.; Coulter, D.E.
EMBL: J. 15, 3182-3196, 1996
A: Title: bowel, an odd-skipped homolog, functions in the terminal pathway during Drosophila
A: Reference number: S70619; MUID: 96272178; PMID: 8670819
A: Accession: S70619
A: Molecule type: mRNA
A: Residues: 1-744 <MAN>
A: Cross-References: EMBL: U08282; NID: g1388165; PIDN: AAB17949.1; PID: g1388166
Genetics:
A: Gene: bow1
A: Cross-References: EMBL: U08282; NID: g1388165; PIDN: AAB17949.1; PID: g1388166
A: Map position: 2L
Keywords: transcription regulation; zinc finger
Query Match 6.1%; Score 322.5; DB 2; Length 744;
Best Local Similarity 18.3%; Pred. No. 3.3e-08;
Matches 195; Conservative 90; Mismatches 281; Indels 501; Gaps 37;
y 90 MDPEHSPPPGSSGPPPTWPERGESSGQFLVA-ATGTAAGGGGLIASPLG-- 146
b 1 MPTESSESSEISGGGGAIFMLRPSR-----MDQFMNSMAAAAVAVGGGLPGCAADRNGS 55

QY 147 -----ATPLPEESTP-----APPPPP----- 164
b 56 GGSDDGSONGNDSDRNSASRISAYETOLAYQOHLAGLHGRPPPPPSHHRETSAPVYL 115
QY 165 ---PPPGVSGHLNIPILILEELRVILQORQIHOQMOETQICQVLLGSLGCTVGAAP 221
b 116 PTGKVRPGSNSNVEIYAMMADKRRELALE-----AAAAAAMLGRRGGGGG 163
QY 222 SELPRTGAASSTKPLPLPFPKIPKAOCTGTASSSSSSSGAEPPKQAFPHLYHPLGSQ 281
b 164 G-VPPPGV-----LGP-----AGVPP----- 180
QY 282 HPFSVGVGRSHKPTPAPSPALPGSTDOLIASPHLAFPGTGTLLAOCIGAARGLEAAAS 341
b 181 -PYLPG-----GSPSTGAS-----SP---FPPGAAMALFPPGLGGMHAGLD 221
QY 342 PGLK- PKNGSGELGYEIVSSLEKPGGRHKCRFCAKVFGSALOIHLSHTEGPERYKC 400
b 222 RLRLRAPGRAS-----RPKQFICKFCNROFTKSYNLIHRTHTDERPYSC 268
QY 401 NYCGRFTTGRNLKVHFRHREKYPHYQMNPHVPEHLDYVITSSGLPYGMSVPEKAEE 460
b 269 DICGAFFRRQDHLRDH-----RLIH----- 288
QY 461 EAGTGGGVGRKPLVASTALSAATESLTLSTGTSTAVAPAGLPTFNKFLMKAVEPKSKA 520
b 289 ----- 288
QY 521 DENTPPGSGSAIAGVADSGSATRMQLSKLVTLSPSMALLTNHLKSTGSPFPYVLEPLG 580
b 289 ----- 288
QY 581 ASPSETSKLOOLVEKIDROGAVAAVASTASGAPTTAPADSSASGPNOCYICLVLSGPR 640
b 289 -----SKERPKCTEGCGFGQSR 307
QY 641 ALRLHYGNGGERPFPKCVCGRAFSTRGNLRAHFVGHKTPSPARAONSCPICOKKFTNAV 700
b 308 TLAVHKLILHMEESPCHKPVCSPRSFNORSNLKTLLTLH-----TDHKYECSGCGVFRNC 363
QY 701 TLOOHVYRMLHGGQIPNGSGA-----LSEGGGA 728
b 364 DLRRLHALTHAVGVNSGDYVDVGEDEARNLGDEEDSLLEVDSPROSPVHNLGSGSG 423
QY 729 QENSSQSTASGSGFPQPOQSPSEEMSEEEDE-----EEEDVDYD 776
b 424 EKSESFRMRKRAAIDHEES-----EEFDDFDEEELQDLRVHDLPREDDDDFPRED 478
QY 777 -----SLAGSGSES-----GGEK-AISVRGDSSEYSG 802
b 479 EQQAEVALVARFOASKAAATSSSSSVGTRKPEROGVTHCHNEGGEYTYMPPHGEKHQDEP 538
QY 803 ABEFVATSVAPPTVTKEMDSNEKAPQHTLPP-----PP-----PPNLDHP----- 844
b 539 GNSGLASLVPVPSPFVR-----YSPRGAAGRPAPRGAPRPPIHQHGHNLPRN 588
QY 845 -----QPMEOGSTSDVSGAMEEAKLEGI-----SSPMALTOEGEGSTPLVEEL 889
b 589 GDEYLPILHVRRLDNLHNSINLSKA-----GVPRPPPTPPTIIORESG--KPLNQL 638
QY 890 NLP-EAMKKDPGSSGSKRACACVCGSFPYQTALEEQKHHPKRGPLFTCYFCRGGFLDRA 948
b 639 HSPHEAMPSTFLGSLPMKRRI-----LPAPTLIDMDPHN--HPGLG-----QRTFVDS 684
QY 949 TLKKHMLAHNOVPPFAPHG--PONTATLTLVPGCSSIPSIPSGLSFP 994
b 685 STIALNMSRH---PPROLGKRPSTETSGATTEKGPVPVAPRIAPRP 728
RESULT 11
B60191
transcription regulatory protein Evi-1, short form - human
N: Alternate names: ecotropic viral integration site 1; oncogene Evi-1

C:Species: Homo sapiens (man)
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 29-Aug-1997
C:Accession: B60191
R:Morishita, K.; Parganas, E.; Douglass, E.C.; Ihle, J.N.
Onogene 5, 963-971, 1990
A:Title: Unique expression of the human Evt-1 gene in an endometrial carcinoma cell line
A:Reference number: A60191; M01D:90326419; PMID:2115646
A:Accession: B60191
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-727 <MOR>
C:Genetics:
A:Gene: GDB:EV11
A:Cross-references: GDB:119889; OMIM:165215
A:Map position: 3q26-3q26
C:Keywords: alternative splicing; DNA binding; zinc finger

Query Match 6.0%; Score 317.5; DB 2; Length 727;
Best Local Similarity 25.7%; Pred. No. 5.5e-08;
Matches 140; Conservative 67; Mismatches 200; Indels 137; Gaps 29;

QY 367 GGRHKCRFCACVFGSDSALQILRS--HTGERPYKCMVCGNFFTRGNLKHFRHREKYP 425
Db 128 GKHYECENCAKVFDPDSNLQHRHSQHVGARAHACPECGKTFATSSGLQKHNHSSVMP 187
QY 426 HVQAMP--HPRVEH-----LDYVITSSG---LPGMSVPP-----EKAEEAGTGGG 468
Db 188 FIFSQSMAPFPDRDRLSRLPKMEPOSPEGVKKLQKGSSESPFDLTTRKRDEKPLTP--- 244
QY 469 VERKPLVASTALSTATESLTLLSTGTSTAVAPGLPTFNKFLMKAVEP-----KSKA 520
Db 245 VPSKRPV--TPATSOQPLD--LSMG--SRSRASG-----TKLTERKKNHVGKKGKS 291
QY 521 DENTPPGSEGSALAGVADSGSAT-----RMQLSKLVTSPLSMALLTNHLKSTGSPFP 572
Db 292 NVESRPASDGS-----LQHARPTPEFMDPIRYVERKRLTD--PLEALKKKYLRSPGFLE 344
QY 573 PYVLEPLGASPE---TSKIQQLVEKIDROGAVAVASGAPTASAPRSSASGPN-- 627
Db 345 ----HPQFOLPDRRTKMSALENMAEKLESFSA--KPEASELLOSVPSEMFRRAPPNAL 397
QY 628 -----OCVLCRLVLSCPRALRL--HYGONGGERPFCKVCVGRAFSTRGNMLRAH 673
Db 398 PENLRLRKGERYTCRVCGKLF--PRSANLTRHLRHTHTGEOPYRCAYCDORSFISSNLQHR 455
QY 674 F--VGHKITSAPAARQNSCPKCKKFTNAVTLQOHVRYMLG-----QIPNGGS 719
Db 456 VNNHNKKEPF-----KCHLCYRCFGOOTNLDRLHKHENGMSGSTATSSPSHELSTGA 510
QY 720 AL-----SEGGAAQENSESOSTASPGSPPOSOQPSPEEESSEE 762
Db 511 ILDDKEDAFETELRNTIGNSNHGSQSPRVNERANGS---HFKEKALVPSQNSDLDDDE 567
QY 763 EDEDEEEDVDDED--SLAGRGSSESGEKAIS-----VRGSEEVYSGAEEVATVAAPT 815
Db 568 EVEDLEVLDEEDDYDITGK--TGKEPVTNLHGNDPEDDEYETSALEMCKTS--PV 621
QY 816 TYKE 819
Db 622 RYKE 625

RESULT 12
JC6510
ras-responsive element binding transcription factor RREB-1 homolog - chicken
C:Species: Gallus gallus (chicken)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: JC6510
R: Miyake, J.H.; Szeto, D.P.; Stumph, W.E.
Gene 202, 177-186, 1997
A:Title: Analysis of the structure and expression of the chicken gene encoding a homolog
A:Reference number: JC6510; M01D:9808435; PMID:9427563
A:Accession: JC6510

```
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1615 <Miy>  
A:Cross-references: GB:AFO13754; NID:g2772826; PIDN:AA96584.1; PID:g2772827
```

	Query Match	6.08;	Score 316;	DB 2;	Length 1615;
	Best Local Similarity	19.84;	Pred. No. 1.5e-07;		
	Matches 234;	Conservative 134;	Mismatches 382;	Indels 434;	Gaps 48;
OY	134 GGGGILLASPLKGLTLPRESTPAPPPPPPPPPGVSGHNLPILEELRVLQOQRQH	193			
Dd	387 GSTGLLSLEPLBATMGPSV-LPTKE-----NIKLL--SLQPPQCFIT	430			
OY	194 OMOMTEQI-----CROVLILSGIQGTGVPASPSELPGTAGASTKPLPLF	240			
Dd	431 QPDSSIVVKRISMSATEIADIQILKNAS-----SAPQIS-LPRLSKAPSV-PVGSTF	483			
OY	241 S--PIKPAQ--TGKTTAASSSSSSSCAEPRPKQAFPHLYHPLGSOHPSPVGVGRSHKP	295			
Dd	484 KHMPLPKPKLVTPRTVAATST-----PP-----PLISAQAASPGCI-----	520			
OY	296 TPASPALPESTDOLI-----ASPHLAFTPTGLLAOCIGANGLEAASPGLLKPK	348			
Dd	521 ---SSLRPPLRLIKNSVETSSNSHSIQRAKSSPSOLL-----LQKVER--LTQH	569			
OY	349 NSGELGYGEVISL-----EKDGR--HKRCPCAK	377			
Dd	570 EMTQLEODSIIEELLPLNMEAKTKQEYTEGDLMAIAGANKKAPTMRKVLYPCRCDQ	629			
OY	378 VFGSDSLQLHLNSHTGERPYKCNC-----GNR-FTR-----GNL	413			
Dd	630 VFASFGLRAHISHLISIPQCNICDIADKAALLIRHLRTHGSERPYICKICHVEGTV	689			
OY	414 KYAHFHRRREYHPNQMPHRPHENDUYITSSGLRYGMSVPRPKAEBAE-----	463			
Dd	690 KANCERLR-KHLKVRKDIEKNIETYTSNAEMVDACFPDYCKLCGEDLKHYHALR	748			
OY	464 ----TPGGGVERKP-----LVASTALSA	483			
Dd	749 IHMRTSGCGKKRFECEKGCTAFSAKNCINHLKHNLHVOREIENYLIVYDCSADES	808			
OY	484 TESLTL-----STGSTAVARGLPTFNKYVL-----	510			
Dd	809 HTDPARLEDSTYMDCKITPRLPEPONGESLGHSSHVPILEPMWGNPMDDEPLDFSQKS	868			
OY	511 --MKAVEPKS-----KADENTPPGEGSAIA-----	534			
Dd	869 KNLSAVOKKENLLVSSPLSFYDCSMERLDLSIPVLYKDKNDLPERGARNOELASSVTIDN	928			
OY	535 -----GVADSGSATRMOL-----SKULTVSPLSMALLTNH--LKSTGSEFP	573			
Dd	929 AYNCCQCPRLFGANNGSEKNRAVGHROPDLKSLHLVPIISPALLGNLSALLRPLRP	988			
OY	574 YLEPLGASPSFSKLQOLVEKIDRGAAVAASTSGAR---TTSAPADSSASAGRQC	629			
Dd	989 ---QPLRPKPVTELPLEA----SIAOIISVSSAPALLKTEADADAPKAASSSTGC	1039			
OY	630 -----VICRLVLSGRAL-----RLHY	646			
Dd	1040 DKSGMAKAKMITYTATIQDSMLPSDLTOACDRPEFIADTGLTKKRKKGTAKPKLISS	1099			
OY	647 G---OHGGERPFCKVCGRAFSTRGNLANHFVGHKTSAPARA-----QN-----	687			
Dd	1100 GVDLSEGE-----FASIEKMLATPTTKRSPLFQSDNFNKESQONCTSDEK	1148			
OY	688 -----SCPICOKKFTNNAVTLQOHVNRHNLGQ---IPNGSAL	721			
Dd	1149 ETPEDKLLRKRNMYSDCPOKITCYPCRFVFWASSILORMHLNHTQSQADTEAPATGGEV	1208			
OY	722 SEBGGAQAQENSSEOSTASGPSFPOQSQOPRPEEMSSEEDDEEP---EED-----	771			
Dd	1209 LIDLTSCEKDEPEEVSELPGSSECPOEEQCADSPRAEDAEDKDEVEEGDEPDSVSNKSL	1268			

[illegible]

Db	615	-----	FFHQIYLKHKHTANCCGAI	SHRRPSPIKHEPADTKDLADIOSILMTSSSS	663	
QY	302	ALPSTQDIIASPH--	-LAPFGTGLLAQCLGAA-----	RG-----	LEAASPGLK 346	
Db	664	SFLRNFGQSVT	PNSSQYSLDGRDDEBAQDAFTSEFRBMKLGRFPCKLCTAVFPLRA		723	
QY	347	PKNSGSLGVEYIS	LSLEKPGRHRCRCAKYFGDSALQIHIRSTGEPYCNCGNR		406	
Db	724	LK-----	GHNRIHLGAVGAPGPRFCNMCYAVCDKAAVLRIHMRITNGDRPYCAVCANYA		777	
QY	407	FTTNGNLKVFH-RRH	-----	RRK-----	YHYVQMP 431	
Db	778	FTTKANCFERHLRNH	GKTSRREYKRAIYYHRAEDACEDCSKSLGDLADTSRSLSTPT		837	
QY	432	HPVP-----	-----	EHLDVYITSS-GLPYGMSVPRPKAE	459	
Db	838	PPRPVNESKSQLK	MLLGENHLPARYNQRP	LKYQKSLDQIVDKKPSAPARQ000000000	897	
QY	460	EEAGT-----	-----	PGGVCERKLVASTAL	SATESLTLSTGSTAYA-----	499
Db	898	EKSGSALDFSM	DVLDLSKKPTPGGASLTPTAVTPTPPAAVAPYTPGCGVGPDLAAALTEOQ		957	
QY	500	-----	-----	PGLPNTKPFYLMKAVEKSKADENTP	-----	525
Db	958	LLLAQDQLFGAG	GYMQDLFSLNFQOSTSGFR	-----	FFPMAAPPPQANPEKPRMVSPP	1013
QY	526	-----	PGSEGAISAGVADSG	SATRMQLSKLYTSLPSMALLTNHLKSTSEFPFVLEPLGA	581	
Db	1014	NRINPMVGVGVG	VPVPVPPGPGVKMVIKNGV-----	LMFKQKQRYRTFRERFACEHSA	1066	
QY	582	SPBETSKLQDL	VEKIDNQ-----	-----	GAVAVASTAGAPRTSAPAPS	620
Db	1067	RFLIRSMERHNV	KQOHPOFYAQRERSAHNVKMGASVNYAAAAAARPMVMAGGPG		1126	
QY	621	SSASGRNQCVIT	GLRVLSCPRALRYHYQ-HG-GERPF	-----	KCKYCGRAFSRGNLRAH	673
Db	1127	SSGFGSNH-----	-----	HGHGHSHGHAIPISDYQKCAILAO-----	OLKAH 1164	
QY	674	FVGHKTSPPARA	ONSCPTCOCKFTNAVTLLQOHVNMHLGQIPN-----	GGSALSEGGA	728	
Db	1165	-----	-----	KNTDILQALAHGSSSVAGNPRLTFGYPLNPSPMHNS--	SQNGGA 1210	
QY	729	-----	-----	QENSSBST-ASGCGSPPOSOOPSREEMSEEBE	-----	DEEF 768
Db	1211	TAMDDEPKLII	DEDENEHDEVAEDVDYF-----	EDEDEDEMDERDEPELLIDEP	1265	
QY	769	EEVDYDEDS-----	-----	LAGRSSEGE-----	KALSVRGDSSEVGAEEVAT	809
Db	1266	AEEAEFEEOGL	PKFLEQGLTKEAQAKMAETLIEQAIKKGKPLSPPTTKENASANTYAT		1325	
QY	810	-----	SVAAFTT-----	-----	VKEMDSNE-KABQHTLPRPPRPDLIDHQ	845
Db	1326	TWQEPATITAT	SVNSSLKMTIAQAEVYKSLKEVASSPDKDSQDVLVPAKLVDMNT	-SQ	1384	
QY	846	PMEQGT-----	SDVSGAMEEAKLEGISSPMAALLOEGBST	STPLVLELNLPRAMKDRCE	901	
Db	1385	NMGFNSYFRPS	DVANAHEMEOSE-DGLVASGASASESNNSGTDEVTSSSSSEPKKAYSL		1443	
QY	902	SSGRKACEVCG	SPPTQALAEHOKTIRKDDPLFCVCRQGFDRATLKKHMLLHHQV		961	
Db	1444	APNKVSCPYC	ORMFPMWSSILRRHLTLTGKRP-FGCSHCPLRLFTTKSCNDRHNRHNV		1502	
RESULT 14						
T13594						
hypothetical protein peb - fruit fly (Drosophila melanogaster)						
C:Species: Drosophila melanogaster						
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999						
C:Accession: T13594						
R:Ferraz, C.; Vidal, S.; Brun, C.; Bucheton, A.; Demaille, J.G.						
submitted to the EMBL Data Library, October 1998						
A:Description: Sequencing the distal x chromosome of Drosophila melanogaster.						
A:Reference number: Z17692						

[illegible]

Map position: X	
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1	
Query Match	
Best Local Similarity 5.7%; Score 300.5; DB 2; Length 2232;	
Matches 242; Conservative 133; Mismatches 453; Indels 327; Gaps 53;	
QY	6 GSSSRILGG--PCGEPAERG-----GDASEHHHPQYCAKCAACPSDPEFLAHONSCCTDP 59
DB	390 GSSSYTGSTTPASSSSSGTMSTMSTGSGTGYVAVPSSSTTGCSSTPIASSSSGSGT--- 446
QY	60 VAVYTG-----GOENPNSA-----SSAPRECHSRQVMDTEHSNPDSC 101
DB	447 VTVVSGSSSTYGSTSPASSSSAGTASTIGSTGSTATIVPSSSSVSGSTQASPSSPG 506
QY	102 S-----SGP-----PDPTWPERGEE-----SSQOFLVAAATGTA 132
DB	507 TMTSTVSGPTGSTVTVVPGSSTSPAPSSSPNPSSSPASTGSTITIGSSSITIVTVSGSTV 566
QY	133 GGGGGL---ILASPKLGATP-----LPPESTPAPPPPPPP-----PPGV 170
DB	567 SSGTGTSGSTLASST--ATPGSSSTVSSSPQSSQSPAPPTGTTPTPSQTSQSFSPSM 624
QY	171 -----GSHLNIPLILEELRYLQQRQIHOMQTEQICQVLLGLSGQTVGAP--A 219
DB	625 NPSSSTPTGSS-----QSTITPEGSTASSPTGSTRFSVATEVT 664
QY	220 SPSELTP-----GTCAASSTPKPLPLFSPIKRQNGKTTAASS--SSSSSGAEP----- 266
DB	665 SOSTVPGSSSLGQSTNTSS---PSPSLSPSTSGMSTLSPSPSSSTGSSQASTLTTP 720
QY	267 ---PQAFPHLYHPL-----GS-----QHPFSYGVGRSHKPPAPPALPGSTDL 311
DB	721 SPNPQSTSLSESTSGATTSAGTMTSPSOSSSYGSSGSGT---SPASTTGEMT 777
QY	312 ASPHLAFPTTGLLAQCLGANGCLEAAAPG--LLKPRKNGSELGYEYVSSLEKPRGR 369
DB	778 SGGSTGTPESSVSTSAAILTSTQOQSVSTNSPGSTVTRPSTVSGSTSSGSTV----- 829
QY	370 HKRCRAKVFSGDSALQIHLSHTGERPKYKNCVGNRTTTRGNLKVHFRHKRYHYQM 429
DB	830 -----VGTSEA-----STSSSVA-----SSSPASTSQ 853
QY	430 NHPVPEHLDYVITSSGLPYGMSVPEPKAESEA--GTPGGVERKPLVASTTALSTAESLT 488
DB	854 NPNPSTSSSSMTQSPYPSQSTSPVSSSTTPSPGSPG-----TTLTSTGSPSQSTT 906
QY	489 LITGTGTAVAFGLPTFNKFLVLMKAVEPKSKADENTPPGEGSAI--AGVAD--SGSA 542
DB	907 IGTGSGST--SPGISTTSE-----EMTSQSGTQT--PGSTGSTVTPQSTVSDSTSSGT 956
QY	543 TRMQLSKLVTS--LPSMALLTNHLKSTGSPFPYVLEPLGASSETSKLOQVEKIDRGA 601
DB	957 VTVGSTEAGSSPLPSTSONTPSTSSS-----SMSTQTPQSSQSTSPV--ESS 1003
QY	602 VAVASTASGAP---TTSAPAPS--SSASGPNQVCLRLVLSGPRALRLHYGCHGERPFK 656
DB	1004 TSGATSSSGSPGTTLSISPSPPSSITIGSSQGSTSPVYVSTISQGSTERPSTGSGTVTKP 1063
QY	657 KVCYGRAFS-----TRGNLRHFV-----GHKTSPARAONSCPICQKFTNAVTL 702
DB	1064 STVSGSASGSTATMGSTEAASSTSGSSTSPNPSTGATSPSTGATSSPSSSGTTLNISIP 1123
QY	703 QOHVNMHLG-----OIP--NGGSALESGGAAGNASSOSTA 738
DB	1124 SPSSQSTISSQGSTSPVYSTGDMTSQGSTQILPGSTGYTQTPSTGSGSTSGEITS 1183
QY	739 SGPSPFPQPOQSPSEEMSEEEDEEEDVDYDSDSLAGRSGESGGEKAISVGDSE 738
DB	1184 QGSTQTPR--SLSTSPAISTSTQOQSVSTNSPGSTVTPQSTV--RGSTSSGSTVYTGTEGS 1241
QY	799 EYSGAHEEATSVAAATPYTKEMDSNEKAPQHTLPPPPPPPNLDRQPMEGTSDVSGAM 858

DB	1242 STSGSSGATSLSSSSPYVPSTSQSPNPSTSGSSTPTPNP-----SQSTSPV----- 1286
QY	859 EEELAKLEGISPMALLOEGGCTSP--LVEELNLPKAKKDPGESS-----GKACAYCG 912
DB	1287 -----VSTTGEWMTSHG--STQTPSTIGSTVTPSTVSGSNSGSGTYVIGSEASTSG 1337
QY	913 QSEPTQALTEHQKTHRKDGLFTFCFCROGFLDRATLKKHMLAHQVFPFAPHQONI 972
DB	1338 SSFKTSPS---SISPVTPSSPIPTTFPASS--TSGSTIS-----DVSSV 1376
QY	973 ATLSLVPGCCSSIPS 987
DB	1377 STTSLAP--LSSSLPS 1390
RESULT 18	
S41688	
DNA-binding protein - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999	
C:Accession: S41688	
R:Fognani, C.; della Valle, G.; Babbiss, L.E.	
EMBO J. 12, 4985-4992, 1993	
A:Title: Repression of adenovirus E1A enhancer activity by a novel zinc finger-contai	
A:Reference number: S41688; MUID:94085372; PMID:8262041	
A:Accession: S41688	
A>Status: preliminary	
A:Molecule type: mRNA	
A:Residues: 1-624 <FOG>	
A:Cross-references: EMBL:X76858; NID:g453618; PIDN:CAA54188.1; PID:g453619	
Query Match	
Best Local Similarity 5.5%; Score 292; DB 2; Length 624;	
Matches 141; Conservative 72; Mismatches 220; Indels 176; Gaps 26;	
QY	337 EAAASP-----GLLKPRKNGSELGYEYVSSLEKGRHKCFCAKVFSGDSALQIHLS 391
DB	3 EAGSPRNHQLGLL-----GEGEQAHVKILYNKE---GRVCMICHHTFKGSLIKAHMVT 55
QY	392 HNGERPCKVNCGRNFTTRGNLKVHFRHKRYPHV-----QMPH-----PV 434
DB	56 HSSRKDHECKLCGASFRTKGLIRHHRHTDERPYKCAKCGKSFRESGALTRHLKSLTPC 115
QY	435 PEHLDYVITSSGLPYGMSVPEPKAEAGTGGGVERKPLVASTTALSTAESLTLLSTGT 494
DB	116 TEKIRPSTSDTAVAGKEEVPAGSSASTVGTYSVAQDMETSPV-----IHLVDAK 168
QY	495 STAVABGLPTFNKFLV--MKAVEPKSKADENTPPGSE-----GSAIAGVADS 539
DB	169 GYIHEVHYVMOEPLRMKMLTPRESPODELPCCSENSRENLLHQAMONGIYLERVAGE 228
QY	540 GSAITRMLSKLVTSLEPMALLTNHLKSTGSPFPYVLEPLGASPSSETSKLOQVEKIDRQ 599
DB	229 ESA-----LEPAPPS-----GSSP-----QCLGDSPLPLIK--VEQIETQ 263
QY	600 GAVAVASTAGAPTSPAPSSASGPNQV-----ICLRVLS-----PRA--- 641
DB	264 -----VASEAATVPRTN--PCQCSSETFPPTAALLKAKRGITAPRPTCTCGCKAPPKAYLL 318
QY	642 -----LRLHYGCHGERPFCKYCGRAFSTRGNLRHF 674
DB	319 KNDQEVHVRFRFGDCGCKLYKTITAHYGRHVRVHSDERFPQPCGKRYKKTNAQOVHF 378
QY	675 VGHKTSPARAONSCPICQKFTNAVTLQOHVNMHLG-----GOIPNG--GSAL 721
DB	379 RTH-----LEEKPHVQDCSRERFKGSLVRHVRHHTGKRPCKYKCGRFPADGTINRL 434
QY	722 SEGGG-----AAQENSSEOST--ASGQSPFPQPOQSPSEEMSEEEDEEED 770
DB	435 KTKGGCLLEVEELLVSESPAAATVLAEDPHTVLVQPSVYADQVEYITTEATDTEFS 494
QY	771 DYTDEDSLGRGSE-----SGGERAI--SVRGDSSEVSGAEKEVA--TS 810

```

b          495 EAT EIEEGTQTEVDSHIMKVVQQLVHQAGAGHQLIVQNTMDQETALGSEAAADIT 552
y          811 VAAPTVAKE 819
b          553 IATPESTLE 561

RESULT 19
01763
trophin-1 - human
Species: Homo sapiens (man)
Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
Accession: G01763
Margolis, R.L.
Submitted to the EMBL Data Library, March 1995
Reference number: G08343
Accession: G01763
Status: preliminary; translated from GB/EMBL/DBD
Molecule type: mRNA
Residues: 1-1184 <MAR>
Genetics:
Gene: GDB:DRPLA; B37
Cross-references: GDB:270336; OMIM:125370
Map position: 12p-12p

Query Match          5.5%; Score 292; DB 2; Length 1184;
Best Local Similarity 21.3%; Pred. No. 1.5e-06;
Matches 236; Conservative 79; Mismatches 399; Indels 392; Gaps

y          42 SDPTETLHAQNSCCDPPWVITIGGQENPNSSSASAPRPGHSRQVMDTEHNPDS- 100
b          118 SDPRDI--DQDNKSTSPSI-YSPGSVENSDSSGLQGGRAPRYNRPPLFPSPQPDST 174
y          101 -----GSSGPRPTWGERERGESSGQFLVA 126
b          175 PRQPEASPEPHNSVYPTGTGYNARMEPPRSKMFQAPRGAPRPRIQLXPGSTGVLGSPRMGP 234
y          127 ATGTAAAGGGG-----LLASPKLGATPLRPBESTP-----APP--- 160
b          235 KGGGAASVVGPRNGKQHNPRTPTISVSSGASGAPPTKPRPTTVGGGGLPSAPPANP 294
y          161 -----PPPGVGS-----GHNITPLIEELKVLQOQRIHQOMTEOI 201
b          295 HTYPLNLPAPRALRPLNNASAPRGLAQRLPGLHLPSP-----HAM----- 334
y          202 CQOVLILGSLGVG-----APASPELPGGAASSTKPLRLPSPKIPRAQNG 249
b          335 -----GCGTGLPQPEKEGPTIAP-SPHLP--PSSSSAPADPMRPFYSSSSSS 380
y          250 KTTASSSSSSSSSGAE--PRKQAFPHLYHRLGSONHPVSGVGGRSHKPTPAAPALPGST 307
b          381 SAAASSSSSSSSSAPFPRAQALPSVPHSPRPPTLSLVS-----SNQPKYTOQSLP--- 432
y          308 DDLIASENILAFPGTTGILLAAQCLGANGRLAASAPGLIKPRNGSELGEGVITSSLEKPG 367
b          433 SOAVWSQGGPRPPRYGLTLA-----NSNAHDPGFPPTGAQ-----STAHNPV 475
y          368 GHHKKEKCAKVFESDALQILNSHMGEPKYKAVCGNRF--TTRGNLKVPHNRHREKXP 425
b          476 STHHNHHDQOQOQOQOQOQHNGNSGRP-----PGAFPHRLLEGSSHHNHPYAMSFS 529
y          426 HVQMNHPV-PEHL-----DYVITSSGLPYGSMVPEKAEKAEAGT----- 464
b          530 LGSLLRPDGPRAILPRPHSQVYSQAQPNQ---PPVSSSSSSSSSTSGGYSYCSHPSPSQ 586
y          465 -PGGVKRPPLVASTALSTESLFTLLST-----GTSTAVAPGLPTNK-----FVLMK 512
b          587 GQGAQVPPRPVPTVTTSSATLS-TVIAIVASSPAGYKTASBPDPGPKRAAPSGQAYKT 645
y          513 AVEPKSKADE-----NTPRGSEGS---ALAGVADSGGATGMOLSKLVTSLPSMALITNH 563
b          646 APPPGKPPSPSPSFRGTGTPGYRGTSPDAGETKPGSP----- 685

```

```

QY 564 LKSTGSPRPVY-----LEBLGASPSFETSKLOQVLEKIDROGAVANASTASGAPTTS 616
Db 686 ---VGPBPLPAGBPSGLPSLPPPPAPASGPPLS-ATQIKOEPAEEXYETPSPPAPAS 740
QY 617 PAPSSASGPNQVCVLCRLVSCPRALRLHYGHONGGERPFCCKVCGRASFSTRENLRAHFVG 676
Db 741 PSPPP-----KYVDYF-----SHASQSAFBNK 762
QY 677 HKTSPAAFAONSCPIQCKKFTNAVTLQOHVMIHGOTPNGSALSEGGAOENSSEOS 736
Db 763 H---LDRGFNSCARSDUYF-----VPLEGSKLAKKRADIVEKVRREA 801
QY 737 TASGPGSPPOQSOQSPREEMSEEEDEDEEDVDDEDLAGRS---ESGGEKAISV 793
Db 802 -----EQARREKEREREREREREKEREKELERSVLAQSGRAVECPSLGVPVH 853
QY 794 RGDSEEVSGAEEVATSVAAPTTVKEMDSNEKARQHTLPPRRPPNDLIDHPOMEQTSO 853
Db 854 RPRPE-PGSAVATVPRLGPDRLATLTLSEVAPRVH-----SGNRHNPVYVLGAVD 906
QY 854 VSGAMEEBAKLEGISTSPMALLQEGBSTPRVIVELNPEAMKQPDGESSGRACEVCGQ 913
Db 907 -----PGLGYNVPAUYSSP----- 922
QY 914 SFPOTALHEHQHTPKDQDPLFTCVFCROGFLDRATLKKHMLDANQV-----PEAPHG 968
Db 923 ---AAREEREAEARND-----LRDLKPRGEYVKPSELEPRLHGVGRGLDPPRHG 969
QY 969 PONTATLSLVPGCSSSTPSPGLSPP 994
Db 970 ---GLALOP-----PPGLMPF 984

```

```

RESULT 20
B32891
finger protein 2, placental - human
C:Species: Homo sapiens (man)
C:Date: 20-Oct-1969 #sequence_revision 20-Oct-1969 #text_change 01-Dec-2000
C:Accession: B32891
R:Bellefroid, E.J.; Lecocq, P.J.; Benhida, A.; Poncellet, D.A.; Belayew, A.; Martial,
DNA 8, 377-387, 1969
A:title: The human genome contains hundreds of genes coding for finger proteins of th
A:Reference number: A32891; MUID:89377476; PMID:2505992
A:Accession: B32891
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-651 <BEL>
A:Cross-references: GB:M27878
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: tandem repeat

Query Match          5.5%; Score 290; DB 2; Length 651;
Best Local Similarity 21.4%; Pred. No. 9.8e-07;
Matches 135; Conservative 63; Mismatches 222; Indels 210; Gaps 20;

QY      364  EKPGGRHKRCFCARVFGSDSALQIHLRSHTGKPYKCNCVNGRTTRGNLKVHFNHR-E 422
        ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      173  EKP-----YNCSQCCKAPSKQSGLTSHQRTHTGKRPYCCEGCCKAFSRKSHLSHWRTHTGE 229

QY      423  KYPHVOMNPNVPVEHLIDYVITSSGLPTGYMSVPEPKAEENEAGTGGGVKRKLVASTITALS 482
        ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      230  K-----PYGCN-----ECG-----RAFS 242

QY      483  ATESILTLSTGTSTAVARGLPTEFNKFLMKAYEPKSKADENTPPSGSGSALAGVADSGSA 542
        : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      243  EKSNI-----NHQRITTEKEPFECRCGKAFRKSQ----- 274

QY      543  TRMOLSKIIVTSIPSMALVTINHLKSTGSFPE-----PVLEPLGASPSSET 586
        ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      275  -----LVTHIRHTGTGKPRFGSCDCKAKAPEKSEBLIRHOTIHTGKRPYC 318

QY      587  SKLGQLVERIKDRGAAVAVASTASCAPTTSAPAPSSSASGENOCYICLRVLVSCPRLRLHY 646

```


406 RFTTRGLTKYHNHNRKRYPHVQMPNRPVRENHDYVITSSGLPRGMSVPRPEAAEEAGRP 465
406 RFTTRGLTKYHNHNRKRYPHVQMPNRPVRENHDYVITSSGLPRGMSVPRPEAAEEAGRP 465
369 TSTTEGRLLKHM---RES-----HTVEQLRAGES-----EPAKSASSP 406
466 GGGVERKPLVASTTALSTALSTESITLLSTGTSTTAVAPGLPTFNKPYLMKAVEPKSKADENTP 525
407 -----KNLSLTKDGSAT--SPINEINLSTTMSI----- 434
526 PSEGSALAGVADSGSATRMQLSKLVTSLPSMALLNLNLKSTGSGFPFPPYVLEPLASPEE 585
435 LDSTNNNAV-----SSTSTPEPSAL-----SALTLDMSST-----PSLSTLAHSSPG 477
586 TSKLQDLVERKIDR-----QGAVAVAS---TASGAPTTASAPSSASGP----- 626
478 VSLALDQIKATISENSFMEPEGINLSALGVSNALIKGDTSPFEKQSNCECRSSSGKITI 537
627 ---NOC-----VICLRVLSCPRA-----LRLHYGNGGERPFKC 657
538 FKCCKGCHGSLSKDDQWMAHARTHIPAERQLOCNCHNCFVTEYKHNLEHYIRNHLIGSKPQC 597
658 KCGRAFTSGNLRAPHVGHKHTSPARAQNSCPICQK----- 694
598 KCGAYNCVKNMSMLNSHMKSHTNHYQFCMD-CTYATKYCHSLKHLKKNHRRVPEGIEM 656
695 -----KETNAVTLQ 703
657 SGGDSPPFTSDATITTSPLMKQELIKETVEPYTSLAQPFPFNMCMGNGLFANHMLLN 716
704 OHVRMLIG---QIPNGGALSSEGGAAQE-----NSSE---QSTVASGSGSPQ 746
717 KHLVDGLMGLRNSVMSPLKCSACDFVASSADEKMRHSMHILNSNVPTSLASLYNSLNL 776
747 PQSQQSPSEEMSEEEDEEBEEDVTDSDSLAGKRSSESGEKALISVRGDEEYVSGAE-- 804
777 PEFSHVAPDNDMALESMDCDVKIDDDNITESHCYEEMDQSDSAVSPTGSSQISSGDEET 836
805 -----EEVATSVAAPTYKEMDS-NEKAPQHTLPRPPPPPNLDHPQMEQGTSDVS 855
837 KKCKSLSLDQISARANGNSPMNSDMSAMEKDEGASADDAIPSPSTTSVSPPLHSSSIYA 896
856 GAMEEFAKLEIGSISSPMALALQEGEGTSPPLVEELNLPEAMKKDPGESSGRKACEVCGSGF 915
897 PIPITPQPNFEFLQSIILAQASLLG---PLL--ANRPSAF-----YCDCKIPIF 938
916 PPTALAEHOKTHPKDPLFTCTVPCROGFLDATALKKMLLAHHQ 960
939 DTQOVLDSHMRFTTPGNP-FWCSDCOYQAFNELSFALHMYQARHQ 982

RESULT 23
31421
-terminal domain-binding protein ral - rat
;Species: Rattus norvegicus (Norway rat)
;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999
;Accession: T31421
;YunyeV, A.; Patturajan, M.; Litingtung, Y.; Joshi, R.V.; Gentile, C.; Gebbara, M.; Cord
;Proc. Natl. Acad. Sci. U.S.A. 93, 6975-6980, 1996
;Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts with
;Reference number: Z21024; MUID:96293459; PMID:8692929
;Accession: T31421
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: mRNA
;Residues: 1-1173 <YUN>
;Cross-references: EMBL:U99056; NID:91438531; PID:91438532; PIDN:AAC52657.1
Experimental source: hippocampus

Query Match	5.4%	Score	286.5	DB	2	Length	1173
Best Local Similarity	20.5%	Pred. NO	2.7e-06				
Matches	257	Conservative	109	Mismatches	418	Indels	467
						Gaps	51

79	PRPEGHSQVMDTEHSNPDGSGSGPPDPFTWGERGEESSCOFLVAATGTAAGGGGL	138

59	VPFSLPLRFAWRKTKTVSPQSHASNP-----ACSRHLILGT-----GDGGPA	102

QY	139	ILASPKIGA---	TLFPRPSTAPRPPPPPPPPGV-----	170
Db	103	PPAPFSSGSSPSSPSSSSSSPPPPPPPPALPAPRFDYDPFHPTDEAVSPPAPE	162	
QY	171	-----GSGHLNTPLLIEELNVLOORQHOMQMTBQICQVLLGSL-----	211	
Db	163	QKTDPFATGSNPSSSGCTPSPDEEED-----EEDDEEGLSOSIRRISETLACTIYDNS	218	
QY	212	-----GQTGAPAPSPSELPGTGAASSTK-----	234	
Db	219	LSQDFPDGDSPHRRPPRPTGLGAPGTPPGADGSTRABGAPRRRVFVVGPEAEACLEGKYSV	278	
QY	235	-----PLPLFSPKPAOTGKTTASSSSSSSSGAEPKOA--FFHLYHPLGSDHF	284	
Db	279	EYVTTAGPPLFL--PLPLP-----TDPEIEGEIYVPEEPRVAVSLFRAAR--ROPPA	330	
QY	285	SVQGVGSHKPT--PAPPALPGSTDLIASHLPALPGTTGLA-----AQCGAAR--	334	
Db	331	SVATLASVAAPAAPAPAPAEGBDFL--SLHADSGALOVLDGPPPPADADAWG	388	
QY	335	-----EAAASPG-----LLKPNXSGSELGYEIVISSLEKPGKRKC	372	
Db	389	GDLRRKILQRRRRYRQRASAPGPPAPKARRERQSRSDPAPD--SPTWEAKKHRS	445	
QY	373	RFCAKVFGSDSAIOILHRSHTGERPKYCNVCGNRTTGNLKVHFHRHREXYPHVQMPH	432	
Db	446	R--ERKJGSHSTARRRSRSSRRRS--RSRSADRR--RGS--HRSRERKRRRRRSAS	496	
QY	433	PVPEHLDYVTTSS-----GLPYGMSVPEKAEEDAG-----TPGGVE	470	
Db	497	PPP-----AASSSSSSRRRRHGRKREGGKKKKKRSRSRREKKSXSGDLEKLPAVPYPSGD	551	
QY	471	R-----KPLVASTALSTATESTLLTGTSTAAPGLPTENKEFYLMAVEBKSK---	519	
Db	552	RDSRRRGAVPSIODLDLHDLEFALKRTTIVGPKDTEPARPS-----PARAVSKGEVLY	606	
QY	520	-----ADENTPGSESGALAGVADSGATRMQLSKLV-----SLPSWALLTNHL	564	
Db	607	DSEGLSADERGAOKDKDRRRRGAASSSSSSREKASRRKALDGDGRDRDSSKKPRPKD	666	
QY	565	KTSGFPFPVLEPLGASPSSETS-----KLQOLVEKIDROGAVANASTASGAPTTSPA	618	
Db	667	SAGSGALPAPRPSRSGSSSSSSCSRRYKLOSKYAVLLREG--VSST-----TPA	715	
QY	619	PSSASGPNQCVICLRYLSCPRAULKHGOHGGERPKVCVCRAVSTRCNLRAHFV---	675	
Db	716	KDSSSSGSLG-----ISVKFSRDEESRPFLEKPD	744	
QY	676	-----GHKTSPPA-----RAQNSCPIQCKKFTNAVTLQOHVHMLGGQIPNGSALSE	723	
Db	745	ERSPAGVAVAPGSTKPKTKAKAKAKAKAGCTGKTKPSKTRK-----KYRSGGSSYAS	800	
QY	724	GG-GAAQEN-----SSDSTASGPGSFPPOQOQSP-----	754	
Db	801	GGGSLSKSKADSCSOAASAKGTEETSWGERTTKAP--STPPKVAAPPALTPDQSOTV	859	
QY	755	-----DEEMSEDEDEDEEDVDVDEDLAGSGSGG	787	
Db	860	DSSCKTPDVSFLAEASBDGVRVGAEEDEEEDEEEDEEEDOPATTTATSTAAAUSTA	919	
QY	788	EKAISVRGDSSEYSGAEAEVATSVAAPTTVKEMDSNEKAPQHTLPPPPPPDNLHDPOPM	847	
Db	920	PSAGSTIAGD--SGAIDGPARAASQJPLT-----PPMPMNL--PAGV	957	
QY	848	EGGTSVSG-----AMEE-----EAKLEIGISSPMAALTQOEGEGTSTPLVEELNLPEAMK	896	
Db	958	DCOTTSVLTALLTALFKMEENLASRAKAOELIQTNOJLHRRKPPSTLGTPAPVPTSFG	1017	
QY	897	KPGGES-----GKKACVCGSPFTQJ-----	919	
Db	1018	LPPAPSSVLLPGSLPTGCGSSTPPTPGCLVPASDKREGSSSEGRGDTDKYLKLHTQER	1077	

[illegible][illegible]

```

119 SDPRDI--DQDNSTSPSI-XSPGSEVENDSDSSSGSQAARPHYHPPPLFPSPQPPST 175
101 -----GSSGPPDPPTWGPERRRGESSGQFLVA 126
176 PROPEASFEHPHSVTPRYNAHMEPPTSRMFOARPGARPNNPOLYPCGTGVLSPGMWP 235
127 ANGTAAAGGGG-----LLASPKLAGPLDEPSP-----APP----- 160
236 KGGGAASVGGPBGKOHPPPTPTISVSSGASGAPPTKPTTPVGGGNLPSAPPANPP 295
161 -----PPPPGVGS-----GHLNIPLLIEELRVLOOROIHOQMOTQOI 201
296 HTYPLNLPPLPALPRLNNAASAPPGAGPLRPHLPSPY----- 333
202 CNOVLLGSLGOTVG-----APASSELPGTGAASSTKPLLPLESPKPAQTG 249
334 -----AMGQMGGLPPEPEKGTTLAP-SPHSLP-----PASSAPAPPMFPYSSSSSS 381
250 KTTASSSSSSSSGAE--PRKQAFPHLHPGLSQHHPFVSGVGRSHKPTAPSPALPST 307
382 SAAASSSSSSSSASAPPSQALPSYHSPPTSLSY-----SNOPKYTOPSLP--- 433
308 DOLIASPHLAFPTTGILLAAQCLGAARGLAAASDGLKPKNGSGELGYGVISLEKPG 367
434 SQAVMSOGPPPPRYGRLLA-----NSNAHGPFPPTGAQ-----STAHPPV 476
368 GHNKCFCAKVFSDSALQIHLRSHTERPYKCNVCNANP--TTKGLKVHFRHREKYP 425
477 STHHHHHQQOQOQOQOQOQOHHNNGSGPP-----PGAFPHPLEGGSSHHAHAYAMSPS 530
426 HVOHNPVY-PEHL-----DYVITSSGLPYGMSVPRPEKAEAEAGT----- 464
531 LGSIRYPPGPAHLPRHSHQVSYQAQPNG---PRVSSSSNSSSTISQGSTPCSHPSQ 587
465 -PGGVEKRRPLVASTALISATLSTLST-----GTSTAVAPGLPTFNK-----FVLMK 512
588 GQGAAPYPPPPVPTVTSATLS-TVIAVASSPAGKYKTASPPGPPRYGKRAPSGAOKT 646
513 AVERPSKABE-----NTPRGSEGS--AIAGVADSGSATRMOLSKVLISLPSWALLTNH 563
647 APPPGYKPSPPSFRGTGPTGKGTSPAGPCTEKGSPPT----- 686
564 LKSTGSFPPPYV-----LEPLGASPESTSKLOOLVEKIDROGAVAVASTAGAPTSA 616
687 ---VGRPLRPAPSGPLSPRPRAAPASGRLS--ATQIKQEPAEYEETPESVPPARS 741
617 PAPSSASGPNOCVICLRVLSCPRALRLHYQONGGERPKCKVCGRASTRGNLRAHFVG 676
742 PSPPP-----KVVDVP-----SHAQSAARNK 763
677 HKTSPARAARONSPICOKKTTMTAVTLOOHVNRHLGGQITNGSGALSEGGAAQNSSESQS 736
764 H---LDRFNSCARSDLYF--VPLE-----GSPWRSQPTWWRRCGAGR----- 802
737 TASGGSFPPOSOQSPSEEMSESEDEEEDVTDSDLAGRS---ESGGEKAISV 793
803 -----AARARRKGAARAAREKEREKEREVERSVKLAQEGRAPVECSLGPVPH 853
794 RGDSEEVGAEEEVATVAAPTYKEMDSNEKARQHTLPPPPPDNLDPHROPMEQSTSD 853
854 RPPFE--PQSAVAVTPPYLGPDPALRTLSEYARBYVM-----SPGNRNHPFYVPLGAVD 906
854 VSGAMEEELKLEISSMALLTQEGEGSTPLVELINLPEAKKKDGESSGKACVCGQ 913
907 -----PGLIGYNPDALYSSD----- 922
914 SFTQTALAEHOKTHPKDGLFTCVCFQGGFLDRATLKKHMLLAHHQVP-----PFAPHG 968
923 ---AARERERARERD-----LRDRLKPGFEVAKPSELEPLHGVPGGLDFFRHHG 969
969 PONTATLSLVPGSSSIPSPGLSPFP 994

```

Db 970 -----GLAIQPG-----PPGLHPPF 984

RESULT 26

S35305

zinc finger protein ZNF91 - human

C:Species: Homo sapiens (man)

C>Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999

C:Accession: S35305

R:Belletroid, E.J.; Marine, J.C.; Ried, T.; Lecocq, P.J.; Riviere, M.; Amemiya, C.; P

EMBO J. 12, 1363-1374, 1993

A:Title: Clustered organization of homologous KRAB zinc-finger genes with enhanced ex

A:Reference number: S35305; MUID:93223677; PMID:8467795

A:Accession: S35305

A:Molecule type: mRNA

A:Residues: 1-1191 <BEL>

A:Cross-references: EMBL:L11672; NID:9186773; PIDN:AAA59469.1; PID:9186774

A:Note: the authors translated the codon GCA for residue 750 as Thr and GCT for resid

C:Keywords: DNA binding; zinc finger

Query Match 5.3% Score 281; DB 2; Length 1191;

Best Local Similarity 21.9%; Pred. No. 5e-06; Mismatches 245; Indels 176; Gaps 22;

Matches 137; Conservative 69; Mismatches 245; Indels 176; Gaps 22;

QY 364 EKPGNRKRCFKAVFSDSALQIHLRSHTERPYKCNVCNANPFTRGMLKVHFRHREK 423

Db 627 EKP---YKCECGKAFSHSALAKHRIHTEKPYKCECGKAFSSSTLANHKTITTEE 683

QY 424 YPHVQNPVPEHNDLYVTSSGLPYGMSVPRPEAE--EACPTGGVEKRRPLVASTAL 481

Db 684 KPYKCECDKTEKRLSTLTNKHITNAG---EKLYKCECG-----KAF 723

QY 482 SATESLTLLSTGSTAVAPGLPTFNKFLVMAKAVEPKSADENTPPGSESGAIGVADSGS 541

Db 724 NRSSNLT-----HKFI--HTGEKPYCEE-----CGKAFNWS 754

QY 542 ATRMQLSKVTSLS-----W-ALLTNMLK-STGSPFPYVLPPLGASPSETSK 588

Db 755 SSLTKHKRIHTEKRPCKCKGKAFIMSSLTLRHKRIHTEG--KPYKCECGKAFSSST 812

QY 589 LQOLVEKIDROGAVAVASTAGAPTSAAPASSASGNOCYTICLRVLSCPRALRLHYGQ 648

Db 813 LTK--HKTITHTG-----EKPYKCECGKAFKHSALAKHII 847

QY 649 HGERPFCKYCGRAESTRGNLRAHFVGHKTSPPARAONSCPICOKKFTNAVTLQOHVM 708

Db 848 HAGEKLYKCECGKAPNOSNLTKHIIHTEKPKSDEE---CDKAFIMSSLTLEKRI 903

QY 709 HLGGQIPNGGSALSEGGAAQNSSEOSTASGP-----GSFPPOSOQSPSEEMSEE 761

Db 904 H-----TREKPYKCECGKAFSQPSHLTTRHMTGEEK 936

QY 762 EEDDEEEDVTDDEBSLAGRSGSGEKAISVRODSEVSGAEEVATVAAPTYKEMD 821

Db 937 PYKCECGKAFSSSTLTTHKIIHTEKPYKC-----EBCGKAFRRSSTLE- 983

QY 822 SNEKAPQHTLPPPPPDNDHPQPMQGSTDVSGAMEEELKLEGISSMAALTQEGEGT 881

Db 984 -----HKIIF-----IGEKPYKCECGKAFSSSTLTTRITRHTTGE 1019

QY 882 STPLVEE---LNLPEAMKDPSESSGRK--ACEVCGSFTQTALAEHOKTHPKDGLP 935

Db 1020 KPYKCECGKAFNRSSKLTHKIIHTEKPYKCECGKAFSSSTLNGKRIHTEKRP-Y 1078

QY 936 TCVFCRQGLDRATLKKHMLLAHHQVP 962

Db 1079 KCECGKAFSSSTLTTRHKLHTEGEP 1105


```

yb 477 IASIAEKYFSTGLVLODKKYGALPYPSMFLPPFPAPFSQSMTPFPDRDLRSLPKMP 536
yy 756 ---EENSSEEBEDEEBEDVT---DEDSLARGSESGBKAIS-----VRGDSSEVS 801
yb 537 QSPSEYKKLÖKSSESPEFDLTTRKRDEKLPTSGPSPGTATPQDDPLDLSMGSRGRAS 596
yy 802 GAE--EEVATSVAAPTGVKEMD---SNEKAPHHTLPPP----- 834
yb 597 GTKLEPRKHNVFEQKKGSNMDTRRSDCSLGARPTFFPMDPTRYVEKKRLTDPLEALK 656
yy 835 ----PPRPNDLRPQMEOGSTDVSGAMEEA-KLEGISS--PMAA-LTGEGESTPIV 886
yb 657 EKYLRSPGFLPHQO-----MSALENMAEKLESFSAIKPRASELLQSVPMSFSFA 707
yy 887 EEELNLPEAMKKDPRESSQRKACEVCGOSPEPTOTALENHONTHPRDGPLETCVCOSGFID 946
yb 708 PENTLPLENLRKGE---RYTCRYCGKITFRPSANLTRHLTLTHTEQP-YRCKYCDRSFYI 763
yy 947 RATLKKNMLLNHOVPPRAH 967
yy 764 SSNLORHVNRINHNKEKPFCH 784

```

RESULT 32
 06546
 finger protein (clone XIC07.1) - African clawed frog (fragment)
 ;Species: Xenopus laevis (African clawed frog)
 ;Date: 28-Feb-1990 #sequence,revision 28-Feb-1990 #text,change 01-Dec-2000
 ;Accession: S06546
 ;Nucleotide, W.; El-Baradi, T.; Mentzel, H.; Pieler, T.; Koester, M.; Poeling, A.; Knoechel, M.
 Mol. Biol. 208, 639-659, 1989
 ;Title: Second-Order repeats in Xenopus laevis finger proteins.
 ;Reference number: S05632; MUID:90040698; PMID:2509712
 ;Accession: S06546

```

;Status: not compared with conceptual translation
;Molecule type: mRNA
;Residues: 1-615 <NIE>
;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
;Keywords: DNA binding; zinc finger

```

Query Match	5.0%;	Score 266.5;	DB 2;	Length 615;
Best Local Similarity	21.6%;	Pred. No. 1.2e-05;		
Matches 135; Conservative	66;	Mismatches 231;	Indels 193;	Gaps 25;

[illegible]

Db 432 CPTNRSQLERHOMITHGEKPISC-PECEECFVSSQLTAHOQOAHRMKPFSCLECGKCF 490

QY 821 DSNERAKPQTLPPRPRLNDHPQMEQGSIDVSGAMEEKEKLEGISSPMALIDEGES 860

Db 491 SNRSNFARHOMITHGEKPFSCSE---CRKGFNSQSLARHQ-----MTHTGE- 534

QY 861 TSTPLVEELNLPBAMKKDPGESSGRKACEVGGSGFPOTALEEHQKTHPKDGLPTVCVC 940

Db 535 ---KPF-----SCSECGKRFRFSNOSHARHOMITHGEKPF-FCSCSEC 570

QY 941 RQGLFDRATLKKHMLLAHNVPPFA 965

Db 571 AKGFNSQSLARHQ-MTHHTGEKPF 594

RESULT 33
526823

zinc finger protein ZNF43 - human
N.Alternate names: zinc finger protein kox27

C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000

C:Accession: S26823; I37967; S10416
R:Löfvering, R.; Trowsdale, J.

Abstract: A gene encoding 22 highly related zinc fingers is expressed in lymphocytes. The gene is located on chromosome 19p13.2 and is transcribed into a 1.5 kb mRNA. The protein is 220 amino acids long and contains 22 zinc finger motifs. The gene is expressed in lymphocytes and is highly related to the zinc finger protein family. The gene is located on chromosome 19p13.2 and is transcribed into a 1.5 kb mRNA. The protein is 220 amino acids long and contains 22 zinc finger motifs. The gene is expressed in lymphocytes and is highly related to the zinc finger protein family.

A:Accession: S26823
A:Reference number: S26823; MUID:912/9444; PMID:1116/5
A:Status: not limited

A:Residues: 1-803 <10V>
A:Molecule type: mRNA
A>Status: preliminary

A: Cross-references: EMBL:X59244; NID:g38031; PIDN:CAA41932.1; PID:g38032
R: Thiesen, H.T.

Title: Multiple genes encoding zinc finger domains are expressed in human

A: Reference number: I37949; MUID: 91145339; PMID: 2288909
A: Accession: I37967

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA

A:Residues: 476-531 <TH1>
A:Cross-references: EMBL:X52358; NID:g34160; PIDN:CAA36584.1; PID:g930090

C,Genetics:
A, Gene: GDB:2NF43; HTF6

A:Cross-references: GDB:128653
A:Map position: 19p13.1-19p12

C;Keywords: DNA binding; zinc finger C;superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homolo

Query Match	5.08;	Score 265;	DB 2;	Length 803;
Best Local Similarity	20.89;	Prod No. 1	00-05.	

Best local similarity 20.08; Freq. no. 1.35-0.5;
Matches 144; Conservative 78; Mismatches 242; Indels 228; Gaps

QY 346 KPNKSGELGYGEVTSLEKPGRHKRCFCAKYFGSDSALQIHLRSHTGERPKCNVCSN 405

Db 287 KAFNÖSSNLT EHKKIRNGEKP --- YKCECGKAFNWPSTLFRHKRIHTGGERPYTCEECGK 343

QY 406 RTTTRGNLKVHFHRH-----REKYPHVQMPHPVE----- 436

Db 344 AFNQFSNLTHKRHTAEKFKYKTECGEAFSRSSNLTKHKIITEKKPPYKCEEGKAFKW 403

```
QY      437 ---HLDVITSSGLPIGMSVPPEEAEEFAGTIPGGVERKPLVASTALSTESLT---LL 490
```

Db 404 SSKTEHRLHTG-----EKPYCEE-----CGKAFNPSTLTKNRI 441

QY 491 STGTSTAVAPGLP-----TENKF-----VLMKAVEPRSKADENTPPGSEGSALGV 536

D6 442 HTGEK-----PYRCEVCGKAFNQFSNLTTHRIHTAEKYCEE-----CGK 483

```
QY 53/ ADGSGANRMOLSKLVTSLPMSWALLNNHLKRSIFPFPPVLEBPGASPSERSKLQQLVEKI 596
```

```

507 DPOGAVUW,ETACCA DMECABADOGCCACCGNVOUTCIVR CGBDRI DFIWCGOYCGNDEBT CEC
DD 484 AFSKSSMELTHKKI-----HIEKK-----PYKCECGKAFKMSKLTIE--HKI 524

```


529 SLTGAASAKDESSPPSGSP--TAGTAATSEHVNOVKATISSVMAAPSTDGMNL----- 580
Y 344 LKPRNGSGELGYGEVYISLEKPGG--RHKCRFCAKVFEGSDSLQILHRSHTGERPKYN 401
Y 581 ---LKNRMWTKYKTLPPYLPKQNGKIKEYCNVCAKTFGOLSNLKLHVRHSGBRPFCCQ 637
Y 402 VCGNFTTGNLKVHFRHREKYPHYQMNPHRVPREHLDVYITSSGLPYGMSVPEKAE 461
Y 638 TCNKGFTQLAHLQKH----- 653
Y 462 AGTPGGVERRKPLVASTLTALSTESLTLLSTGTSTAVAGLPFTFNKFLMKAVEPKSKAD 521
Y 654 -----LVHGC----- 658
Y 522 ENTTPGSEGSALAGVADSGSATRMQLSLVTSLPSSWALLTNHLKSTGSPFPVYLEPLGA 581
Y 659 ----- 658
Y 582 SPSESTKLOQLVKEKIDRGAVAVASTAGAPTTSAAPSSASAGPNOCVICLRVLSCPRA 641
Y 659 -----BK-----PHECQVCHKRPSSTSN 676
Y 642 LRLHYGOHGERPFCKVCGRAFSTRGNLRAHFVGHKTPAARAQNSCPTCQKKFTNAVT 701
Y 677 LKTHLRLHSGEKRYQCKVCAKFTQFVHLKL---HKRLHTRRPHKCAQCHKYIHLS 732
Y 702 LQOHVRLHGGQIPNGSALSSEGGAQENSSQSTASGPGSPFPOQSPREEMSE 761
Y 733 L---KVHLKGNCPAGPAA-----GLPLEDLTRI 757
Y 762 EEEDE-----EEEDVTDSDSLAGRSESGEKAI--SVRGDSEFVS 801
Y 758 NEIERFIDISDNADRLMEDSDVDTSMV-----EKETLAVVRKEKETS 802

RESULT 36

inc finger protein 51 - mouse
;Alternate names: finger protein zfc12
;Species: Mus musculus (house mouse)
;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 08-Dec-2000
;Accession: I48668; A40984
;Buike, P.S.; Don, J.; Wolgemuth, D.J.
fam. Genome 5, 387-389, 1994
;Title: Zfp-51, a murine zinc finger encoding gene mapping to the t-complex region of C
;Reference number: I48668; MUID:94319090; PMID:8043957
;Accession: I48668
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: mRNA
;Residues: 1-710 <RES>
;Cross-references: EMBL:X74855; NID:9488832; PIDN:CAA52847.1; PID:9488833
;Cross-sley, P.H.; Little, P.F.R.
;Title: A cluster of related zinc finger protein genes is deleted in the mouse embryo
;Reference number: A40984; MUID:91376058; PMID:1680234
;Accession: A40984
;Status: preliminary
;Molecule type: mRNA
;Residues: 672-710 <CRO>
;Cross-references: GB:M74235
;Genetics:
;Gene: zfp-51
;Superfamily: zinc finger protein zfp-36; LIM metal-binding repeat homology
;Keywords: DNA binding, zinc finger

Query Match 5.0%; Score 262.5; DB 2; Length 710;
Best Local Similarity 21.8%; Pred. No. 2.2e-05;
Matches 144; Conservative 64; Mismatches 265; Indels 189; Gaps 22;
Y 364 EKPGRRHKCFCAKVPFSDSLQILHRSHTGERPKYCNVCGNFTTGNLKVHFRHREK 423
Y 140 EKP---HQCCKGCKFMTSSSLTVHHRITKPKYCSVCDKSFCTQCTHLKIHORRHTE 196

QY 424 YPHVQMNPHRVPREHLDVYITSSGLPYGMSVPEKAEBAAGTPGGVERKPLVASTTALSA 483
Db 197 KP-----KCECESEFYQLSA 213
QY 484 TESTLTLSTG-----TSTAVAGLPFTFNKFLMKAVEPKSKADENTPPGSESAI 533
Db 214 LKSHQKLTGKRPYKCECDKSFANHPNFRTHQKI-----HTSEHSCDCEGREF 264
QY 534 AGVADSGSATRM-----QLSKLVTSLPWALLTNHLKSTGSPFPVYLEPLGASPSFT 586
Db 265 HOLSHLRKHRYLHTGERPKYCNCDRSFTTHYASLRWHQ--THTPEIHCKEBCGKSTEL 323
QY 587 SKLOO-----LVEKIDRGAVAVASTAGAPTTSAAPSSASAG--PNOCVICTL 633
Db 324 SHLKHHRYHTGKRPYKCECDK-----STTTTLTKHQKHNGKRYKCECD 373
QY 634 RVLSCPRLRLHYGOHGERPFCKVCGRAFSTRGNLRAH---FVGHKTPAARAQNS-- 688
Db 374 KSFHSHLRHRYHTGKRPYKCECDKSFESATLREHKSNGEKTYKCECDKSF 433
QY 689 -----CPTCQKKFTNAVTLQOHVRLHGGQIPNGSALSSEGGA 728
Db 434 QRAYLRNHNHRYHTGERPKYKCEGKSFCTCTYHHTGK---PYCTEGCKSF 489
QY 729 QENS---SEOSTASGPGSPFPOQSPREEMSEEEDEEDVTDSDSLAGRSES 785
Db 490 TNNSTLRTHQKHS-----EKPRCKGCDKSFSGSTLKAHQSIH 530
QY 786 GGE---KAISVRGDSSEVS--GAEEVATSVAAPTTVKEMDSNEKAPDHTLPPPPPPDN 840
Db 531 TGEKPYKIECDKSFQVSHLRTHQRYHTG--ERPRCTECDKSFIRSSH----- 578
QY 841 LDHPQMEGTSIVSGAMEEAEKLEGISPMALQLQEGSGTTPVLELNLPEAMKDPG 900
Db 579 LREHQKHS-----EKPYKCRDCCDISFSQIS-----NLRHQKLTG 616
QY 901 ESSGRKACEVCGSPFQTALAEHQKTHPKDGLPTCYFCRQGLDRAVLKIMHLAHQ 960
Db 617 EKPRK--CMEDCKSFTHNSFRTHQRYHTGKPR--YRCAECDSFTTCSYLRARHQKHTGE 673
QY 961 VP 962
Db 674 KP 675

RESULT 37

finger protein glaas - fruit fly (Drosophila melanogaster)
;Species: Drosophila melanogaster
;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 21-Jul-2000
;Accession: S05447
;Moses, K.; Ellis, M.C.; Rubin, G.M.
Nature 340, 531-536, 1989
;Title: The glaas gene encodes a zinc-finger protein required by Drosophila photorec
;Reference number: S05447; MUID:89365138; PMID:2770860
;Accession: S05447
;Molecule type: DNA
;Residues: 1-604 <MOS>
;Cross-references: EMBL:X15400; NID:98015; PIDN:CAA33450.1; PID:98016
;Note: 60-Gln, 361-Ser, 362-Leu, and 377-Thr were also found
;Genetics:
;Gene: glaas
;Cross-references: FlyBase:Fbgn0004618
;Map position: 3R 91A1-2
;Introns: 16/1, 59/3; 557/2
;Keywords: DNA binding; nucleus; transcription factor; zinc finger
;439-571/Region: zinc fingers

Query Match 4.9%; Score 262; DB 2; Length 604;
Best Local Similarity 19.3%; Pred. No. 1.9e-05;
Matches 125; Conservative 48; Mismatches 171; Indels 304; Gaps 18;

```
OY 118 ESSGQFLVAATGTAAGCGGLIASPKLIGATPLPEESTAPRPPPPPP----- 168
:| | : : | | : | | : | | | | | | | |
221 QSSGNNTGSSNSA-----GGVANASCYET---SAGTAAPRRPAAAMYPMSVAVSMN 263
OY 169 -----GVSGCHL-NIPDLIELRLVLOQROIHOMQTEQICROVLLLSLGQTVGAPASPS 222
:| | : | | : | | : | | : | | : | | : | |
284 MTMHNGYGGDAGGVDM-----QCSQMWT-----PPSNS 313
OY 223 ELPGHGAASSTKPLLLPLFIKPAQTKTTASSSSSSSSSSSGAEPKQAFPHLYHPLGSOH 282
Db 314 TSAAMAAAVNVLLPYPLSP-----GHYPASATYSPFADFRAPAPTG-----LGALP 360
OY 283 PFSVGVGGRSHKPTPAAPSPAL-----PGSTDQLASPHLAFPGTGLLAQCSLGAARGLE 337
:| | : | | : | | : | | : | | : | | : | |
361 PMT-----VGEKESPSPANSLSLGYPTPGVGNQGYTPRHKS-----PTSYQAALLGLS 409
OY 338 AAASGGLKPKNGSGELGVEGVISLSLEKPGGRHK---CRFCAKVFSGDSALQIHLRSHHG 394
:| | : : : : | | : | | : | | : | | : | |
410 LSA-----FDEDEDSNEDLDGDEGSS---GGEKMPNLCLGCKTYARPSTLKTHLRTHSG 461
OY 395 EEPYCNVCGNRTTGNLKVHFHHRBEKYPHYQMNPRVPHNDLYVITSSGLPYGMSVP 454
Db 462 EEPYKCPDCKSKFSQAANLTAHVRTH----- 487
OY 455 PEKAEAEAGTGGGVBERKPLVASTTALSTELTLSTGTSTAVAGLPTEFNKFLMKAV 514
Db 488 ----- 487
OY 515 EPKSADEMTPPGSESSATAGVADSGSATRMQSLKLVTSLSWMLLTNHLKSTGSFPFPPY 574
Db 488 -----TGQKPF-- 493
OY 575 VLEPLGASPSETSKLOOLYVEKIDROGAVAVASTASGAPTTSAPAPSSSASGPNQVCICLR 634
Db 494 -----RCPIICDR 500
OY 635 VLSCPRALRLHYGQHGGERPFKCKVCGRASFTSGNLRAHFVGHKTSPPARAONSFCPICOK 694
Db 501 RFSQSSSVYTHMHTHGERPYRCCSKKFSFDSSTLTGHLRTH-----SEKPYQCACLCL 556
OY 695 KFTNAVTLQOYHVMHLGCGQTPNGSGALSSEGGAOENSSEOSTASGPG 742
Db 557 RFSQSGNLRLHMKVHGNNNNSNGSNGATGVGG-----ESSTGSGVG 597

RESULT 38
S06548
finger protein (clone X1COF8.4) - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 01-Dec-2000
C:Accession: S06548
R:Nietfeld, W.; El-Baradi, T.; Mentzel, H.; Pleier, T.; Koester, M.; Poeling, A.; Knoech
J. Mol. Biol. 208, 639-659, 1989
A:Title: Second-order repeats in Xenopus laevis finger proteins.
A:Reference number: S05632; MUID:90040698; PMID:2509712
A:Accession: S06548
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-536 <NTE>
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match 4.9%; Score 261.5; DB 2; Length 536;
Best Local Similarity 19.9%; Pred. No. 1.8e-05;
Matches 119; Conservative 62; Mismatches 215; Indels 203; Gaps 20;

OY 372 CRFCAKVFSGDSALQIHLRSHTERPYKCNVGNRFTTGNLKVHFHHRREKYPHYQMP 431
:| | : | | : | | : | | : | | : | | : | |
Db 8 CSECKCFAGSELNVRHRTHTKVPSCSGCKFCSNQTLYKHRTGTGKRP----- 61

OY 432 HPRVPHLDVYITSSGLPYGMSVPEKAEAEAGTGGGVBERKPLVASTALSTESTLTLIS 491
Db 62 -----FSCSECKCFCS 72
```

```
OY 492 TGTSTAVAPGLPTFNKRVLMKAVEPKSKADENTPPGSE--GSALAGVADSSATRMQSLK 550
:| | : | | : | | : | | : | | : | | : | |
Db 73 TPR-----VAARHOKTHTGKPFPCSECKCFARSSD----- 104
OY 551 VTSLSWMLLTNHLKSTGSFPFVLEPLGASPSFETSKLOOLYVEKIDROGAVAVASTASG 610
:| | : | | : | | : | | : | | : | | : | |
Db 105 -----VTIV-RRTHTGKPYSCSGCKCFETSSSDL-----NVHRR----- 138
OY 611 APTTSAPAPSSASGPNQVCICLRYLSCPRALRLHYGQHGGERPFKCKVCGRASFTSGNL 670
Db 139 -----THTGKPYSCSHCGKCFETSSSELNVHRTHTGKPYSCSECKGFPTSEF 189
OY 671 RAHFVGHKTPARAONSCPTCKKFTNAVTLQOYHVMHLGGOIPNGGSLSGEGGAOE 730
:| | : | | : | | : | | : | | : | | : | |
Db 190 TSHWKTTHMEKPF---SCVOGCKCFSKDTHLKYHRTHTGK---PFCFECGKCFTH 241
OY 731 NSS-----EOSTASGPGSPPOQSPPEEEMSEEEDEEEDVTDSDSLAGR 781
Db 242 NGSLLKYLHKIKREADCCKGNLTNPEI---SPVEHYPTTNEIKEE-----AT 286
OY 782 GSESGEKAISVNGDSEEVSGAEEVATSVAAPTTVKEMDSNEKADQTLPPRPPDNL 841
Db 287 SMEGNOQSDYSINSLFQIQGP-----YTPTIMEYN-----HLI----- 321
OY 842 DHPQPMQGTSDVSGANDEEAKLEGISSPMALPDGEGSTPIVLE-----ELNPEAMK 896
Db 322 ---MODNKYDVNAC-----HSPVPL-----OETDYLKALHNRDIDRQRTQOTOK 362
OY 897 KDPGESSGRK--ACEVCGSPPTOTALNEHOKTHPKDGLFTCVFCROGLDRAITLKKH 953
Db 363 YDHRTNTGDKRLPSCSECKCFSTYHVLARHOKTHTGKRF--FCSCECKCYARSSDLNVH 420

RESULT 39
S51037
zinc-finger protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 01-Dec-2000
C:Accession: S51037; S06565
R:Schaefer, U.; Rausch, O.; Boumeester, T.; Pleier, T.
Eur. J. Biochem. 226, 567-576, 1994
A:Title: Sequence-specific recognition of a repetitive DNA element by a C(2)H(2) zinc
A:Reference number: S51037; MUID:50094815; PMID:8001572
A:Accession: S51037
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-675 <SCH>
R:Nietfeld, W.; El-Baradi, T.; Mentzel, H.; Pleier, T.; Koester, M.; Poeling, A.; Kno
J. Mol. Biol. 208, 639-659, 1989
A:Title: Second-order repeats in Xenopus laevis finger proteins.
A:Reference number: S05632; MUID:90040698; PMID:2509712
A:Accession: S06565
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 85-448, 'K', 450-613 <NTE>
A:Experimental source: clone X1CGF20-1
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match 4.9%; Score 261.5; DB 2; Length 675;
Best Local Similarity 20.1%; Pred. No. 2.3e-05;
Matches 128; Conservative 75; Mismatches 255; Indels 179; Gaps 21;

OY 364 EKPGRHNRKCRFCAKVFSGDSALQIHLRSHTERPYKCNVGNRFTTGNLKVHFHHRREK 423
:| | : | | : | | : | | : | | : | | : | |
Db 143 EKP---YSCQCGCKSFANHCVLDSHORHTGDDKPPSCTECGKFKFSQGNLHKHLKTHKLD 199
OY 424 YPH-----YQMNPRVPHNDLYVITSSGLPYGMSVPEKAEAEAG 463
Db 200 OPHLCAECGKTSFSPKSTLLEHOKTHSELTPSEKFTSDANHLKHOSTTE----- 252
OY 464 TPGGVBERKPLVASTA-LSTESTLTLSTGTSTAVAGLPTEFNKFLMKAVEPKSKADE 522
```


Db 883 KAPSPSETCDSEISEAP--VSPEWAPPDGAESGGKPPSPSQVQOOSYHTQPRLVHQHNI 940
QY 868 SSPMALIQEGEGTSTPL-----VEELNLP---EAMKKDPGE----- 901
Db 941 QVPEIRVTEEPDPPEKEKEKAQSKPEKPEVEEFOWPQORSETLSQLPAEKLPPKKRLRLAD 1000
QY 902 ---SSGRKACEVCGQSF---PTOTALEEHOKTHPKDGPLFTCVFCROGFLDRATLKKHML 955
Db 1001 MEHSSGESSEFESTGTGLSRSPSQESNLSSHSS-----FSMSFEREETSKLSALPKODE 1053
QY 956 LAHQVPPFAPHPGPNITATLSLVPG-----CSS-SIPSPGLSPFP 994
Db 1054 FGKHSEFLTVPAGSYSL-----VPGHHHQKEMRCCSSBQMPCPHPAEVP 1098

Search completed: January 13, 2003, 15:21:29
Job time : 54.4664 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:17:40 : Search time 11.4828 Seconds
(without alignments)
3619.260 Million cell updates/sec

Title: US-09-988-117-3
Perfect score: 5301
Sequence: 1 MAOETGSSRLGPGCEPAE.....SSIPSPGLSPFPKDDPTMP 1002

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5182	97.8	1004	1	SAI2_MOUSE
2	4618.5	87.1	1007	1	SAI2_HUMAN
3	1338	25.2	1300	1	SAI3_HUMAN
4	1301	24.5	1324	1	SAI1_HUMAN
5	1282	24.2	1323	1	SAI3_MOUSE
6	1273.5	24.0	1322	1	SALI_MOUSE
7	759.5	14.3	1355	1	SALM_DROME
8	711.2	13.4	1402	1	SALM_DROME
9	371.5	7.0	1845	1	SALM_DROVI
10	323.5	6.1	773	1	Z236_HUMAN
11	311	5.9	734	1	ZN42_HUMAN
12	306.5	5.8	654	1	Z133_HUMAN
13	303	5.7	722	1	Z219_HUMAN
14	301	5.7	1829	1	Y296_HUMAN
15	290	5.5	738	1	ZN64_HUMAN
16	290	5.5	1185	1	DRPL_HUMAN
17	290	5.4	1350	1	XFTN_XENLA
18	287.5	5.4	982	1	HB1L_CAEEL
19	283.5	5.3	2688	1	ZEP1_MOUSE
20	282.5	5.3	818	1	KR18_HUMAN
21	281	5.3	1191	1	ZN91_HUMAN
22	278.5	5.3	512	1	Z169_HUMAN
23	274	5.2	706	1	Z151_CHICK
24	271.5	5.1	898	1	Z071_XENLA
25	271	5.1	1060	1	ZF01_DROME
26	270.5	5.1	780	1	Z084_XENLA
27	270.5	5.1	839	1	ZF95_HUMAN
28	270	5.1	882	1	ZN45_HUMAN
29	269	5.1	1042	1	EV11_MOUSE
30	268.5	5.1	717	1	PRDD_HUMAN
31	268	5.1	697	1	Y441_HUMAN
32	267	5.0	578	1	Z192_HUMAN
33	267	5.0	683	1	Z263_HUMAN

34	266	5.0	845	1	KR0H_DROME	P08155 drosophila
35	265	5.0	803	1	ZN43_HUMAN	P17038 homo sapien
36	263.5	5.0	686	1	ZN07_HUMAN	P17097 homo sapien
37	263	5.0	803	1	Z151_HUMAN	O13105 homo sapien
38	263	5.0	1167	1	Z208_HUMAN	O43345 homo sapien
39	262	4.9	604	1	GLAS_DROME	P13360 drosophila
40	261.5	4.9	645	1	ZF93_MOUSE	O61116 mus musculu
41	261.5	4.9	759	1	Z287_MOUSE	O9e9b9 mus musculu
42	261	4.9	428	1	ZF26_MOUSE	P10076 mus musculu
43	260.5	4.9	1342	1	Z335_HUMAN	O9h422 homo sapien
44	260	4.9	594	1	ZF37_MOUSE	P17141 mus musculu
45	260	4.9	751	1	Z184_HUMAN	O99676 homo sapien
46	259.5	4.9	803	1	Z226_HUMAN	O9h716 homo sapien
47	259.5	4.9	1183	1	DRPL_RAT	P54258 rattus norv
48	259	4.9	553	1	Z324_HUMAN	O75467 homo sapien
49	258.5	4.9	751	1	Z337_HUMAN	O9y3m9 homo sapien
50	258.5	4.9	794	1	Z151_MOUSE	O60821 mus musculu
51	257.5	4.9	675	1	ZG20_XENLA	P18714 xenopus lae
52	256	4.8	453	1	Z06_XENLA	P18749 xenopus lae
53	256	4.8	2161	1	SHK1_HUMAN	O9y566 homo sapien
54	255.5	4.8	720	1	PRDH_HUMAN	O9h9d4 homo sapien
55	254.5	4.8	595	1	ZN85_HUMAN	O03923 homo sapien
56	254.5	4.8	754	1	Z287_HUMAN	O9hbt7 homo sapien
57	254	4.8	439	1	Z028_XENLA	P18747 xenopus lae
58	252.5	4.8	670	1	ZN16_HUMAN	P17020 homo sapien
59	252	4.8	2167	1	SHK1_RAT	O9ww48 rattus norv
60	251.5	4.7	1163	1	Y222_HUMAN	O92618 homo sapien
61	251.5	4.7	1258	1	GLI2_HUMAN	P10070 homo sapien
62	250.5	4.7	477	1	MAZ_HUMAN	P36270 homo sapien
63	249	4.7	616	1	ZN93_HUMAN	P35789 homo sapien
64	249	4.7	739	1	YD49_HUMAN	O9p2j8 homo sapien
65	248	4.7	476	1	Y106_HUMAN	O96119 homo sapien
66	248	4.7	595	1	Z317_HUMAN	O96p67 homo sapien
67	247.5	4.7	477	1	MAZ_MOUSE	P56671 mus musculu
68	247.5	4.7	783	1	ZF25_HUMAN	O9u1i5 homo sapien
69	247	4.7	892	1	HIC1_MOUSE	O9-rt1y mus musculu
70	247	4.7	947	1	Z268_HUMAN	O14587 homo sapien
71	246.5	4.7	913	1	Z228_HUMAN	O9u1u3 homo sapien
72	246.5	4.7	1051	1	EV11_HUMAN	O03112 homo sapien
73	246	4.6	589	1	ZN44_HUMAN	P15621 homo sapien
74	245.5	4.6	469	1	Z135_HUMAN	P52742 homo sapien
75	245.5	4.6	755	1	RRE1_HUMAN	O92766 homo sapien
76	244	4.6	510	1	ZF64_HUMAN	P15622 homo sapien
77	244	4.6	614	1	ZF29_MOUSE	O07230 mus musculu
78	244	4.6	821	1	ZN41_HUMAN	P51814 homo sapien
79	244	4.6	1276	1	PRDC_HUMAN	O9ha22 homo sapien
80	244	4.6	2717	1	ZEP1_HUMAN	P18622 homo sapien
81	243	4.6	582	1	ZF36_HUMAN	P16415 homo sapien
82	242	4.6	488	1	Z345_HUMAN	O14585 homo sapien
83	241.5	4.6	354	1	KLF2_MOUSE	O60843 mus musculu
84	241	4.5	698	1	Z234_HUMAN	O14588 homo sapien
85	240	4.5	351	1	KLF2_RAT	O9e158 rattus norv
86	240	4.5	457	1	Z140_HUMAN	P52738 homo sapien
87	240	4.5	627	1	Z264_HUMAN	O43296 homo sapien
88	239.5	4.5	688	1	HKR3_HUMAN	P10074 homo sapien
89	239	4.5	509	1	ZKRI_CHICK	P30373 gallus gall
90	239	4.5	636	1	ZF90_MOUSE	O61967 mus musculu
91	239	4.5	819	1	ZF95_MOUSE	O9z1d8 mus musculu
92	239	4.5	1687	1	Z142_HUMAN	P52746 homo sapien
93	238.5	4.5	803	1	ZXDB_HUMAN	P98169 homo sapien
94	238.5	4.5	3866	1	HRX_MOUSE	P55200 mus musculu
95	238	4.5	601	1	ZF37_RAT	O88553 rattus norv
96	237.5	4.5	733	1	HIC1_HUMAN	O14526 homo sapien
97	237	4.5	1029	1	Z197_HUMAN	O14709 homo sapien
98	237	4.5	1403	1	CIC_DROME	O9u1h0 drosophila
99	237	4.5	3703	1	ABF1_HUMAN	O15911 homo sapien
100	236.5	4.5	614	1	ZF28_MOUSE	P10078 mus musculu

ALIGNMENTS

RESULT 1

[illegible]

QY	196	QMTEDQICROYLLLLSGQTVGAPASBEBLPTGTAASTRKPLPLPSPIKPAOTGKTTASS	2555
Db	198	QMTEDQICROYLLLLSGQTVGAPASBEBLPTGTAASTRKPLPLPSPIKPAOTGKTTASS	257
QY	256	SSSSSSGAEPPKQAFPHLYKPLGSGDPFVSQGVGRSHKPRPAPSPALPGSTDLTASPH	315
Db	258	SSSSSSGAEPPKQAFPHLYKPLGSGDPFVSQGVGRSHKPRPAPSPALPGSTDLTASPH	317
QY	316	IAFPGTTGLAAOCLGAARGLAAASFGLLKPKNGSGELGYEIVISSLEKPGGRHKCRFC	375
Db	318	IAFPGTTGLAAOCLGAARGLAAASFGLLKPKNGSGELGYEIVISSLEKPGGRHKCRFC	377
QY	376	AKVFGSDSLQIHLRSHGTGERPKKCNVCGNFTFRGNLKVHHRHREKXPVHYOMNHPVP	435
Db	378	AKVFGSDSLQIHLRSHGTGERPKKCNVCGNFTFRGNLKVHHRHREKXPVHYOMNHPVP	437
QY	436	EHLDDYITSSGLDYPGMSVPEPKAEAEAGTPGGYERKPLVASTALSAATESLTLSTGTS	495
Db	438	EHLDDYITSSGLDYPGMSVPEPKAEAEAGTPGGYERKPLVASTALSAATESLTLSTGTS	497
QY	496	TAVAPGLPFFNKFLMKAVBPCKSKADENTPPGSBGSATAGVADSASATMQLSKLVTSLP	555
Db	498	TAVAPGLPFFNKFLMKAVBPCKSKADENTPPGSBGSATAGVADSASATMQLSKLVTSLP	557
QY	556	SMALLTNHKKSTGSPFPFPPYLEPLGASPSERSKLOOLYEKTIIDROGAVAASTASGAPTTSS	615
Db	558	SMALLTNHKKSTGSPFPFPPYLEPLGASPSERSKLOOLYEKTIIDROGAVAASTASGAPTTSS	617
QY	616	APAPSSASGPNOCVITCLRVLSCPRALRLHYGHGHEGRRPKCKVCGRAPSTRGNLRAHFV	675
Db	618	APAPSSASGPNOCVITCLRVLSCPRALRLHYGHGHEGRRPKCKVCGRAPSTRGNLRAHFV	677
QY	676	GKHTSPAAARQNSCPTCQKKFTNAVTTLQOHHBMHGGQIPNGSALSBEGGAQENSSQ	735
Db	678	GKHTSPAAARQNSCPTCQKKFTNAVTTLQOHHBMHGGQIPNGSALSBEGGAQENSSQ	737
QY	736	STASGSGSPPOPSOQPPSPBEEMSEEEDEEEDVDDEDLSDAGSGSESSEGGKATSVRG	795
Db	738	STASGSGSPPOPSOQPPSPBEEMSEEEDEEEDVDDEDLSDAGSGSESSEGGKATSVRG	797
QY	796	DSEEVSGAEEVATVSAAPTVTKEMDSNEKAPQHTLPPRRPPDNLDPHPMEQSTSDVS	855
Db	798	DSEEVSGAEEVATVSAAPTVTKEMDSNEKAPQHTLPPRRPPDNLDPHPMEQSTSDVS	857
QY	856	GAMEEPAKLEGISSPMAALTOEGEGSTPLVEELNLPRAMKDDPGSSGGKACEVGOGSF	915
Db	858	GAMEEPAKLEGISSPMAALTOEGEGSTPLVEELNLPRAMKDDPGSSGGKACEVGOGSF	917
QY	916	PQOTALHEHQTKPRKQGLPTCYFCRCRGFLDRATLKKHMLLANHNOVPRPAPHQNIATL	975
Db	918	PQOTALHEHQTKPRKQGLPTCYFCRCRGFLDRATLKKHMLLANHNOVPRPAPHQNIATL	977
QY	976	SLVPGCSSSIPSPGLSPFPKKDDPTMP 1002	
Db	978	SLVPGCSSSIPSPGLSPFPKKDDPTMP 1004	
RESULT 2			
SAL2_HUMAN			
ID	SAL2_HUMAN	STANDARD;	PRT; 1007 AA.
AC	O9Y467; O9Y4G1;		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Sal-like protein 2 (Zinc finger protein SAL2) (Hsal2).		
GN	SAL2 OR KIA0360.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_Taxid=9606;		
NP	[1]		
SEQUENCE FROM N.A. (ISOFORM 1).			

AL3 HUMAN STANDARD; PRT; 1300 AA.

Q9BXA9; Q9UGH1; 15-JUN-2002 (Rel. 41, Created)
15-JUN-2002 (Rel. 41, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Sal-like protein 3 (zinc finger protein SALL3) (hsSALL3).
SALL3.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 3).
Gooding R., Angelicheva D., Blechschmidt K., Svoboda K., Molnar M.,
Tournev I., Kalaydjieva L.;
"Exclusion of HSALL3 and refinement of the region for the CCFDN
gene."
Submitted (Feb-2001) to the EMBL/Genbank/DBJ databases.
[2]
SEQUENCE OF 29-1300 FROM N.A., AND ALTERNATIVE SPLICING.
MEDLINE=20079154; PubMed=10610715;
Kohlhase J., Hausmann S., Stojmenovic G., Dickens C., Bink K.,
Schulz-Schaeffer W., Altman M., Engel W.;
"SALL3, a new member of the human spalt-like gene family, maps to
18q23."
Genomics 62:216-222(1999).
-1- FUNCTION: Probable transcription factor.
-1- SUBCELLULAR LOCATION: Nuclear (Probable).
-1- ALTERNATIVE PRODUCTS: 3 isoforms; 1, 2, 3 (shown here) and 4; are
produced by alternative splicing. Isoforms 1 and 2 lack two zinc
finger domains and are the major isoforms.
-1- TISSUE SPECIFICITY: Widely expressed in adult with highest levels
in heart. Expressed in fetal brain (in neurons of hippocampus,
cortex, mediodorsal and ventrolateral thalamic nuclei, putamen,
cerebellum and brainstem).
-1- DEVELOPMENTAL STAGE: In fetal brain of the 24th gestational week.
-1- SIMILARITY: BELONGS TO THE SALL FAMILY OF C2H2-TYPE ZINC-FINGER
PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF347021; AAK18311.1; -
EMBL; AJ007421; CAB65124.1; -
HSSP; P07248; IARE.
GeneW: HGNC:10527; SALL3.
MIM: 605079; -
InterPro: IPR000822; znf_C2H2.
Pfam: PF00096; zf_C2H2; 10.
SMART: SM00355; znf_C2H2; 10.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.
PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
Metal-binding; Repeat; Alternative splicing.
Zn_FING 420 442 C2H2-TYPE.
Zn_FING 448 470 C2H2-TYPE.
Zn_FING 679 701 C2H2-TYPE.
Zn_FING 707 729 C2H2-TYPE.
Zn_FING 739 761 C2H2-TYPE.
Zn_FING 977 999 C2H2-TYPE.
Zn_FING 1005 1027 C2H2-TYPE.
Zn_FING 1113 1135 C2H2-TYPE.
Zn_FING 1141 1163 C2H2-TYPE.
DOMAIN 198 201 POLY-ALA.
DOMAIN 213 216 POLY-GLN.
DOMAIN 897 902 POLY-SER.
VARSPPLIC 1 133 MISSING (IN ISOFORM 2 AND ISOFORM 4).

FT VARSPPLIC 973 1044 MISSING (IN ISOFORM 1 AND ISOFORM 2).
FT CONFLICT 235 235 R -> C (IN REF. 2).
FT CONFLICT 593 593 V -> L (IN REF. 2).
FT CONFLICT 787 787 D -> N (IN REF. 2).
FT CONFLICT 797 802 DDMDE -> NDNDLK (IN REF. 2).
FT CONFLICT 808 808 D -> N (IN REF. 2).
FT CONFLICT 1138 1138 F -> K (IN REF. 2).
FT CONFLICT 1141 1141 F -> S (IN REF. 2).
SQ SEQUENCE 1300 AA; 135371 MW; 82CF3BDCB8D59150 CRC64;
Query Match 25.2%; Score 1338; DB 1; Length 1300;
Best local similarity 32.2%; Pred. No. 6.2e-51;
Matches 377; Conservative 117; Mismatches 342; Indels 334; Gaps 40;
Db 4 ETGSSRLGGPCGEPARERGDASEHHHPQVCAKCAQPSDFLEFLAHQNSCCDPPMYI 63
38 DSGPESRSGG-----EETSVKCCAEFFKMADFLEHQRKSTKLPLVLI 82
64 ICGENPSSSASAPRPEGHSKQVMDTEHNPDSGSGPPDPWGERGESESGQF 123
83 -----HEDADAP-----RHEDPE-----PSPASSEERASEEAE- 113
Qy 124 LVAATGAAGGGGGLIASP-----KLGATPLPESTPPAPPPPPPGVSGHLN- 176
114 -----EAGAEAGEGEAREVEKEAEEMDAEPAGDTRAPRPPAPAPPTPAYGAPSTNV 166
Qy 177 -----IPILFELRYLQOQHOMQTEOI 201
Db 167 TLEALLSTVAAVAFSOGARAGSGAGGVAAAVPRLLEDGLMALQOQDIDLDLEOI 226
Qy 202 CROVLLLG-----SLGQTVGAP-SPSELPCTGAASSTRPLPLFSPKPAOTGKT 252
Db 227 RSQVALMQRRPPRPSLSAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 266
Qy 253 ASSSSSSSSGAPRPOAFPHLYHPL-----GSOHPFVGVGRSHKPPAPAPALD- 304
Db 267 SAGAPAAALAGSGPAPAPAEAGAPLSPRESGASTPPGPAEBSAPAPAPAPAPAPA 326
Qy 305 -----GSTDOLIASPHLAFPGTGLLAQCIGARGLEAASP----- 342
Db 327 PAPQSAASPOASATPPALAPGS-----LLGAPGLPSLLPOTSAGVIFPNLYS 379
Qy 343 -----GLKPKNGS-GEIYGGEVYISLEKPGGRHKCRFCAYVGSALDIDL 389
Db 380 IAAFTANALDPLSLMKHRRKKPRPVSVFEPRKASADPFKHKCRFCAYVGSALDIDL 439
Qy 390 RSHTEBRPKCNVCGNRETTGNLKVPHRREKTPHYOMNPNVPHLDVYITSSGLPY 449
Db 440 RSHTEBRPKCNVCGNRETTGNLKVPHRREKTPHYOMNPNVPHLDVYITSSGLPY 499
Qy 450 GMSVPPER-----AEEAGTPGCGVERKPLVA 476
Db 500 GMSLPPERKVTYTLDSKPYLPYPTSVGLQPLPYTPGAGHYADSPATRASRPQRSPA 559
Qy 477 STTALATFSLTLLSTGTS-TAAVAP-----GLPTNKFFYLKMAVEPKSKADENTPPGSEGS 531
Db 560 SSECASLSPGLNIVESGVSAFESQSLGGRPVTK-----AEVSLPCTNARAD- 610
Qy 532 AIAGYADSSATRMQSLKLVSLPSMALLTNHLKSTGSPFPFYVLEPLGASPESTSLQ 591
Db 611 --APVGAQASAPRTSVGAPTSLSGSLPLPAVSEQFAKPFGLDLSM--QTSETSKLQ 666
Qy 592 LVEKIDROGAVAAVAVASGAPPTSAPAPSSASGNOVCYICRLVSLCPALRLHYGHHG 651
Db 667 LVENIDKR-----MTDPNOCYICRVLSCQSAALKKHYRTHTG 703
Qy 652 ERPFCKVCGRAFSTRGNLRAHFVGHKTSAPAARQNSCPICOKKFTNAVTLQOHVRLHG 711
Db 704 ERPFCKICIGRAFTTGNLKTGHGVHAKRPLRLVQHSCPICQKFTNAVTLQOHVRLHM 763
Qy 712 GOIPNGSALSBEGGAQEN--SSEOSTASGPGSFPQPOSQSPSEENKSEEBDEDEEE 769
Db 764 GOIPN--FLPREGFQDAMSELAYDKNAETLSY-----DDMDENSMEDDAEL 811

2N_FING 1134 1156 C2H2-TYPE.
 2N_FING 1162 1184 C2H2-TYPE.
 DOMAIN 150 159 POLY-SER.
 DOMAIN 160 163 POLY-SER.
 DOMAIN 237 240 POLY-SER.
 DOMAIN 294 297 POLY-SER.
 DOMAIN 371 375 POLY-SER.
 DOMAIN 1144 1147 POLY-SER.
 VARIANT 150 150 /FTID=VAR_013155.
 VARIANT 150 150 /FTID=VAR_013156.
 VARIANT 159 159 /FTID=VAR_013157.
 VARIANT 164 164 /FTID=VAR_013158.
 VARIANT 1265 1265 /FTID=VAR_013159.
 VARIANT 1265 1265 /FTID=VAR_013159.
 CONFLICT 79 79 G -> A (IN REF. 1: CAB41399).
 SEQUENCE 1324 AA; 140390 MW; CBE787847A96D8AA CRC64;

Query Match 24.5%; Score 1301; DB 1; Length 1324;
 Best Local Similarity 31.3%; Pred. No. 2.5e-49;
 Matches 383; Conservative 142; Mismatches 338; Indels 360; Gaps 48;

4 ETGSSRLGPPC--GEPAERGGDAEEHHROYCAKCAQFSDPTERFLAHQNSCCTDPPM 61
 17 EVASLPRRRGDTEKGGPS-RPTKSKDAH--VCGRCACFEFLSDLLHKKKCTKNQVL 72
 62 VIIGGOENPSSASSAPRP-----EGHSRSQVMDTEH 94
 73 IV---NENGSPETSPSPRPNDQMDVNTKTDQVDCDLSHNGDRESMEVEA 129
 95 --SNPPDSSGGPPDPTWGPERRGESSGQFLVAATGTAAGGGGILLSPKLGATPLPP 152
 130 PVANKSGSGTSS-----GSHSTAPSSSSSSSGSGSGTGTSAITTSILPQ 178
 153 -----ESTP-APPPPPPPPPPPVSGSHNIPILIELRYLQO 189
 179 LGDLTTLGNFVINSVNIINLSTKYAVAOFSQARCGGASGKLAVPALMEQLALQO 238
 190 ROIHOMQTEQICROYLLGLSGTVGAPASPSLPGT----- 227
 239 QOIHQLOLEQIHNQILLAS--QNAOLPSSPSGQGLRTSANPLSTLSSHLSQLAA 296
 228 -----GAASSTKLPLLPSP-----IKPAQGTGK-----TTASSSS 258
 297 AGIAGSLASQASISGVKOLPPIQPOSSGNTIIPNSGSSPNMILAAAVTPPSSK 356
 259 SSSSGAEPKQAFHLYHPL---GSHPFVSGVGRSHKPTAPSPALP--GSTQDLIAS 313
 357 ASSAGAS-----HNSNPVSSSSSPAFAT-----SLLSPASNPLLPQOASANSVPPS 404
 314 PHLAEPPTGLLAQCGAARGLBAASPLGLPKNGSGELGYEYISLEKPGCRHKCR 373
 405 P-LPNTGT---AEDLNS---LSALAOQRKSKPENVTA---FEAKSTDEAFKHKCR 452
 374 FCAYFGSALQIHLSSHGERPYKVCNRRPTKGNLKVHFRHREKYPHYOMNHP 433
 453 FCAYFGSALQIHLSSHGERPYKVCNRRPTKGNLKVHFRHREKYPHYOMNHP 512
 434 VPEHLDYITSSGLPYGMSVPEKAEAEAGTPGGVERKPLVASTALASSTLSTG 493
 513 VPEHLDNIPSTGTGPIYMSIPE-----KPV---TSLMDKPVLPPLTTS 554
 494 TSTAVALPGLTFENKFLAKAVEKS---KADENTPPGSEGSALAGVADSG---SATR--- 544
 555 VGLPLPPLSLPLPI---KTEEPAPIPISHSATSPGSVKS-----DSGPEESATNMG 606
 545 -----MQLSKIYV-SLP-----SMALLTNTL----- 564
 607 GLPEAEAGSTLPPSGKSESGMNTNSVPTASSSVLSVSPADCGPAGSATITFTNPLFLM 666

QY 565 --KSTGSEPPYVLEPLGASPESTSKLOOLVEKIDROGAVAVASTAGAPTTSAPAPSS 622
 Db 667 SEQKAFPFEGGLD--SAQASFTSKLOOLEVENDKK----- 701
 QY 623 ASGNOCVITCLRVISCPRALRLHYGONGGERPKCYGCAAFSTRGNLRAHFVGHKTS 682
 Db 702 ATPNECITICRVLSQCSQALKNHYRTHTGERPKCKICGRAFTKGNLKVHFRHREK 761
 QY 683 ARAONSCPICQKFTNATVLOOHVHRMLGGQIPN-----GGSALSSGGAAOEN--- 731
 Db 762 LRYHSCPICQKFTNATVLOOHVHRMLGGQIPN-----GGSALSSGGAAOEN--- 731
 QY 732 ---SSEGTASGPGSFPQ--POGQOPSPPE-----EMSE----- 760
 Db 822 DNFEDENMEDPEGSIPDTPKADASQDLSSSPFLPMSIALLENQMKMINAGAEOL 881
 QY 761 ---EEEDDEEEDVDDBSLAGRSESGEKAISYRGDSSEVSGAEEVATVAAPT 816
 Db 882 QASLSYVNGSLGDLVLTNDS-----SSVGGMESQAGSPAPSESTSSMOA 928
 QY 817 VKEMDSNE---KAPQHTLPPPPPPNLDH---POPMEOGTSVSGAMEEAKLEGISSP 870
 Db 929 LSPNSITQEFHKSPTIEKQRAVPSSEFANGLSFTPVNGALDLTSSRAK----- 979
 QY 871 MAALTOEGEGTSTPLVEELNLEPMARKDPGESSGRKACEVCGSPPTOTALHEHOKTHP 930
 Db 931 DGPLTYCFRCROGLFDRALTKMLAHQV-----PPAPGPQNIATLSVPGCS 982
 Db 980 --LIKEDSLGLFPF-----RDGKFK-NFACDIDCGKTFACQALDIHYRSHTK 1025
 QY 1026 ERP-FICTVNCNRFSTKGNLKHQMLT--HQMRLPSQLFEPSSNLGNONSAVTPANSL 1082
 Db 983 SSIPSP-----GLSPFRKDDPT 1000
 Db 1083 SLIRTEVNGFVHVSPODSKDTPT 1105

RESULT 5
 SAL3 MOUSE
 ID SAL3 MOUSE STANDARD: PRT; 1323 AA.
 AC Q62255;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sal1-like protein 3 (Spalt-like protein 3) (Msal) (Fragment).
 GN SAL3 OR SAL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Embryo;
 RX MEDLINE=96391179; PubMed=8798152;
 RA Ott T., Kaestner K.H., Monaghan A.P., Schuetz G.;
 RT "The mouse homolog of the region specific homeotic gene spalt of Drosophila is expressed in the developing nervous system and in mesoderm-derived structures.";
 RL Mech. Dev. 56:117-128(1996).
 CC -1- FUNCTION: Probable transcription factor.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are produced by alternative splicing. Isoform 2 lacks two zinc finger domains (6 and 7) and is the major isoform.
 CC -1- TISSUE SPECIFICITY: In adult brain, testis and kidney. In lower levels also in adult ovaries and embryonic stem cells. In embryo in developing neuroectoderm of brain, inner ear and spinal chord. Also weakly and transiently expressed in embryonic branchial arches, notochord, limb buds and heart.
 CC -1- DEVELOPMENTAL STAGE: During embryogenesis detected from 7 dpc onward in tissues derived from mesoderm and ectoderm.
 CC -1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X97581; CAA66196.1; -
 DR HSSP; P07248; IARE.
 DR MGD; MGI:109295; Sal13.
 DR InterPro: IPR000822; ZnF_C2H2.
 DR Pfam; PF00967; zf-C2H2; 10.
 DR SMART; SM00355; ZnF_C2H2; 9.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
 KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
 KM Metal-binding; Repeat; Alternative splicing.
 FT NON_TER 1 1
 FT ZN_FING 398 420 C2H2-TYPE.
 FT ZN_FING 426 448 C2H2-TYPE.
 FT ZN_FING 663 685 C2H2-TYPE.
 FT ZN_FING 691 713 C2H2-TYPE.
 FT ZN_FING 723 745 C2H2-TYPE.
 FT ZN_FING 968 990 C2H2-TYPE.
 FT ZN_FING 996 1018 C2H2-TYPE.
 FT ZN_FING 1104 1126 C2H2-TYPE.
 FT ZN_FING 1132 1154 C2H2-TYPE.
 FT DOMAIN 119 123 POLY-PRO.
 FT DOMAIN 188 191 POLY-GLN.
 FT DOMAIN 345 348 POLY-SER.
 FT DOMAIN 881 886 POLY-SER.
 FT VARSPIC 964 1035 MISSING (IN ISOFORM 2).
 FT SEQUENCE 1323 AA; 139070 MW; 86083DEAEF2CE2D0D CRC64;
 SO
 Query Match 24.2%; Score 1282; DB 1; Length 1323;
 Best Local Similarity 31.3%; Pred. 1.66-48;
 Matches 374; Conservative 133; Mismatches 329; Indels 360; Gaps 45;
 QY 4 ETGSSRLGGPCGPAERGGDASEHHPOVCAKCAQAFSDPTFLAHQNSCTDPYAVT 63
 DB 9 DSGSSRSRG-----SEE--TSVCEKCAEFKFMADFQHKKTKCTKNPLVLI 53
 QY 64 IG-----GQENPSNSASS-APRPEGHSRSQVMDTHS-----NPPDSSGSGR 105
 DB 54 HDDEAPRPSSEDPPEPSRASSPSDRTESEVAEEVALEPTESEYKAAATKEASMDVEVSTDGK 113
 QY 106 PDPPTWGPERRKEESSGQFLVAATGTAAAGGGGGLILASPKLATPLRPESTPAPRPPPP 165
 DB 114 PGP-----GVP-----SVPPPPPALPQ 128
 QY 166 PPP-----GVGS--GHNLPLLIEELRVLQ 188
 DB 129 PEPAFASMPSTNVLTETLLSTKYVAVAQFSOGARAGCTTGAGSGVGAVALPMLIEQLALQ 188
 QY 189 QROIHQMTQEQICROVLLGSGQTVGAAPSPLE-----P 225
 DB 189 QOOIHQOLITEIRSY---GLMSRQGPRLPKFASAPGNTSVOLGTLPHNALQLSAGR 245
 QY 226 GTGAASSTRKPLPLFSPPIKPAQGTGKTTASSSSSSSSGAERPKQAFNHLNPLGQNP-- 283
 DB 246 ATASGSGSTPLRAAFD--GPQHLISQPASGTSPPCSTSAAPRDSG--AHFACSTGPAP 299
 QY 284 ----FVSVGVGGRHKR-----TP-----ASPALPGSITDQLASPHLAFPGT 321
 DB 300 GAVAAASSTVGAADVQMASTPRLALGPGPLLSASNLDPNLPQT-----SSSVLTPNP 354
 QY 322 TGLLAQCLGAARGLEMAASPGILPKNGS--GELGIVGVISLLEKPGGNHCKRFCAKVF 380
 DB 355 LVSTIAA-----TANALDPLSA--LMKHKKGRPNVSVFERPKASAEDEPFKKKCFCAKVF 408
 QY 381 SDSALQILHRLSHTEGRPYKCNVCGNRTTQGNLKVHPIHNRKITYRNVQNNPVRPHLDY 440

Db 409 SDSLALILHRSHTGERPFCKNICGNFRFSTKGNLKVHFQHKKEYPHIQNNPVYPLELDK 468
 Qy 441 VITSSGLPRGMSVPPER-----AEEEGATP-----GGYERKPLVASTNLSATESLTLS 491
 Db 469 CPTSGSIPYGMSTLPPEKRVTTWMLDSKVPVLPIVPMVSVDGLRPETVGTNYTSDSPSTPPVS 538
 Qy 492 -----TGSTAAVAGLPTEFKFYLMKAVERK-----SKADEMTPGSE---GS 531
 Db 529 RSPQRSPASSSETSLSPGLNNTESGITVRRESQPRLGSPSLTAEPVSLPCTSTRICD 588
 Qy 532 A-IAGYADSG---SATRMQLSKLVTSI--PSWALTINHLKSTGSPFPYVLEPLGASPSE 585
 Db 589 APVVGGQVAVGGLPRTSAATAVYDASACTSGSPCLPAVSDQPKA--QEPFGGLDSQM--QTSE 644
 Qy 586 TSKLOOLVEKIDMGAVVAVASTASGAAPTTSAPARSSASGPNOCYCIALNVLSCPRALRLH 645
 Db 645 TSKLOOLVENIDKK-----MTDPNQCYICHANVLSQSGLKMH 681
 Qy 646 YGOHGGERPFCKVCVGRAFSTRCNLRAHFVGHKTSPARAONSCPICOKFTNAVTLQOH 705
 Db 682 YRTHTGERPFCKXICGRAFTTKGNLKFHVGRKPRLRVQHSCPIDCQKFTNAVYLQOH 741
 Qy 706 VMHNLGGQIPN-----GGSALSFCGGCAOENSSQO-----ST 737
 Db 742 IIMHMGGQIPNPRLPEGLOEAMADLPDEFKENAETLSFSDDIDENSMEDSELKDTASD 801
 Qy 738 ASGP-----GSFPQPOSQOPSPREEEMSE-----EEDEDEEE 769
 Db 802 SKKPILLSGSGCP-----PSPREVISIALLENOMKMIDSVMNOCOLANLKSVENGSGE 855
 Qy 770 EDYTDSDSLAGR----SESGEKAIIVRGDSEEVSGAEAEVATSVAAPTTVKEMDSNEKA 826
 Db 856 SDRLSNDSNSAWADLESRSAGSPALSSSSSQALSPANSNGESPFSKSPGLGHOEDPOEI 915
 Qy 827 POHT--LPPPPRPDNLDHPQMDQIGSDVSGAMEEAKLEISSPMALVTQEGESTSP 884
 Db 916 PLKTERLDSPPEPGN-----GCALDLTGTGHC-----RP 945
 Qy 885 LVEELNLPEA-----MKKDPSGESSGKACBVCOSQSFPTOTALAEHQKTPKPDGLFTCV 938
 Db 946 LIKE----EAFSLFLFSRGRKCAS-TVCGVCCGFACKSALTEHYRSHTEK-RRFVCT 999
 Qy 939 FCRGRTLDRAATLKKMHLAHQVPRFAPHGQNATLSLYVGCSSSIPIBPGLSPFP 994
 Db 1000 VCRGCGSTMGNLKHÖHLT--HKLEL---PSQVFDPNFRTLGPSHTSPSLASSPAR 1049

RESULT 6
 SALL_MOUSE STANDARD; PRT: 1322 AA.
 AC Q9ER74; G92OR5;
 DT 15-JUN-2002 (Rel. 41, Created)
 RX 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sal-like protein 1 (Zinc finger protein Spalt-3) (Sal-3) (Msal-3).
 GN SALL1 OR SAL3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/Ola;
 RX MEDLINE=20422467; PubMed=10965108;
 RA Buck A., Archangelo U., Dixkens C., Kohlhasse J.;
 RT "Molecular cloning, chromosomal localization, and expression of the
 RL murine SALL1 ortholog Sal1-1.";
 RL Cytogenet. Cell Genet. 89:150-153(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21543523; PubMed=11688560;
 RA Nishitakamura R., Matsumoto Y., Nakao K., Nakamura K., Sato A.,

Copeland N.G., Gilbert D.J., Jenkins N.A., Scully S., Lacey D.L., Katsuki M., Asashima M., Yokota T.;
"Murine homolog of SAL1 is essential for ureteric bud invasion in kidney development.";
Development 128:3105-3115(2001).
[3]
FUNCTION, AND ASSOCIATION WITH HDAC.
MEDLINE-21964044: PubMed-11836251.
Kiefer S.M., McDill B.W., Yang J., Rauchman M.;
"Murine sal1 represses transcription by recruiting a histone deacetylase complex.";
J. Biol. Chem. 277:14869-14876(2002).
-1- FUNCTION: transcriptional repressor involved in organogenesis. Essential for ureteric bud invasion in kidney development. Homozygous deletion of SAL1 results in an incomplete ureteric bud outgrowth, a failure of tubule formation in the mesenchyme and an apoptosis of the mesenchyme.
-1- SUBUNIT: Interacts with HDAC1, HDAC2, RBBP4, RBBP7, MTA1 and MTA2. Probably associates with NURD histone deacetylase complex (HDAC).
-1- SUBCELLULAR LOCATION: Nuclear.
-1- TISSUE SPECIFICITY: Expressed in the metanephric mesenchyme surrounding ureteric bud.
-1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

	EMBL: AJ271914; CAC09602.1; -. EMBL: AJ271915; CAC09602.1; JOINED. EMBL: AB051409; BAB5673.1; -. HSSP: P15822; IBOO. MGD: MGI:1889585; Sal11. InterPro: IPR000822; Znf_C2H2. Pfam: PF00096; Zf-C2H2; 10. ProDom: PD000003; Znf_C2H2; 1. SMART: SM00355; Znf_C2H2; 9. PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9. PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9. Transcription regulation; Repressor; DNA-binding; Nuclear protein; Zinc-finger; Metal-binding; Repeat. Zn_FING 450 472 C2H2-TYPE. Zn_FING 478 500 C2H2-TYPE. Zn_FING 705 727 C2H2-TYPE. Zn_FING 733 755 C2H2-TYPE. Zn_FING 765 787 C2H2-TYPE. Zn_FING 1000 1022 C2H2-TYPE. Zn_FING 1028 1050 C2H2-TYPE. Zn_FING 1133 1155 C2H2-TYPE. Zn_FING 1161 1183 C2H2-TYPE. DOMAIN 133 139 POLY-THR. DOMAIN 237 240 POLY-GLN. DOMAIN 1143 1146 POLY-SER. CONFLICT 165 165 C -> S (IN REF. 2). CONFLICT 167 167 S -> T (IN REF. 2). CONFLICT 1271 1271 S -> SS (IN REF. 2). SEQUENCE 1322 AA: 140229 MW: 448477 PFA0395FC C6C64.
--	---

Query Match	24.0%	Score 1273.5;	DB 1;	Length 1322;
Best Local Similarity	31.6%	Pred. No. 3.7e-48;		
Matches 377:	Conservative 144;	Mismatches 372;	Indels 299;	Gaps 48;

RESULT 7
ID SALM_DROME STANDARD: PRT: 1355 AA.
AC P39770: 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeotic protein spalt-major.
SALM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94139659; PubMed=7905822;
RA Kuehnlein R.P., Frommer G., Friedrich M., Gonzalez-Gaitan M.,
Weber A., Wagner-Bertholz J.F., Gehring W.J., Jaechle H., Schuh R.,
"Spalt encodes an evolutionarily conserved zinc finger protein of
novel structure which provides homeotic gene function in the head and
tail region of the Drosophila embryo."
RL EMOB J. 13:168-179(1994).
CC -1- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF THE POSTERIOR-MOST
CC HEAD AND THE ANTERIOR-MOST TAIL SEGMENTS OF THE EMBRYO. PROBABLY
CC FUNCTION AS A TRANSCRIPTIONAL REGULATOR. COULD REPRESS THE
CC TRANSCRIPTION OF THE TSH GENE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT BLASTODERM STAGE AND
CC LATER IN RESTRICTED AREAS OF THE EMBRYONIC NERVOUS SYSTEM AS
CC WELL AS IN THE DEVELOPING TRACHEA.
CC -1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X75541; CAA53229.1; -
DR HSP: P15822; LBAO.
DR FlyBase: FBgn004579; salm.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 7.
DR PRINTS: PR00048; ZINCfinger.
DR SMART: SM00355; Znf_C2H2; 7.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
KW Developmental protein; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Transcription regulation; Repeat.
FT ZN_FING 451 473 C2H2-TYPE.
FT ZN_FING 479 501 C2H2-TYPE.
FT ZN_FING 824 846 C2H2-TYPE.
FT ZN_FING 852 874 C2H2-TYPE.
FT ZN_FING 884 906 C2H2-TYPE.
FT ZN_FING 1289 1311 C2H2-TYPE.
FT ZN_FING 1317 1339 C2H2-TYPE.
SQ SEQUENCE 1355 AA; 148995 MW; 6FC4EBDC0BC6355E CRC64;
Query Match 14.3%; Score 759.5; DB 1; Length 1355;
Best Local Similarity 22.7%; Pred. No. 5.2e-26;
Matches 312; Conservative 141; Mismatches 415; Indels 505; Gaps 44;
3 OGTGSSRRIGGPGCEPAERGDASEEHHPVCAKCAQCFSDTEFLAHNSCTDPPVMY 62
Db 50 KIGSGQENGSGSPLTTATTTASPSRSP-----PEEOPPEOSTSQST---PEOST 100
OY 63 IIGGQENPNSASAPRPGHSHRQVMDTEHSN---PPDSGSSGPPDPTWGPERGEES 119

Db 101 PDHLENDIKSEAKSEIEPEDNNNNVAMTKPSSSEEREPNAGSMSSPV--AEASAE 158
OY 120 SGQFLVAATGCTAAGGGGLIASPKLGATP-----LPPESTPAPPPPPPP 166
Db 159 ATE-----RTPEKEKEKVEVDVEMDEAPSSAVPSTEVTL 194
OY 167 PGVSGSHLNPILLEELR-----VIG 188
Db 195 PGAGA-----PVTLLEIQNMQALIAQFAATTIANGSGNADAMKQLAFLQOTLENLQ 249
OY 189 QROIHOMQMTQICROYLL-----LGSIGQTGA 217
Db 250 QOOLFQIQLQLOSLALNQAKQEDTEDADQEQDEQETDYEEERIADIMELRQKA 309
OY 218 PASPSLPGGAASSTPLPLPSPIRKAOIGKTTASSSSSSSSGAEPPKQAFPHLYR 277
Db 310 EARMAE-----AKARQHLINAGVPLR-----ESSGPAESLKRREHDH- 348
OY 278 LGSQHPFVSGVGRSHKPTPA-----PSPALPGSTD--QLTASPH--LAPQTTG 323
Db 349 --ESQPNRRTSLDNTHKADTRADALAKLKEMENTPLPGSDLAISITTHNDLPERNSLD 406
OY 324 LLAAGCIGARGLEAASPGLK-----PRNGSELGYEVISSLEKPGRRKCR 373
Db 407 LQKR-----AQEVLDSASQGITLANSMAADPFAFGKSGEKK-----GRNEPFFKHRCR 454
OY 374 FCAGVGSDSALQIHLRSHGERPKYKCVGCGNRFTRGKLAKYHFRHREKRYPVQNNPHR 433
Db 455 YCGKVGSDSALQIHLRSHGERPKYKCVGCGNRFTRGKLAKYHFRHREKRYPVQNNPHR 514
OY 434 VPEH-----LDVYTTSSGLPYGMSVPRPEKAEAGTGG--GVER--KPLVASTAL 481
Db 515 IPEHMKFHPRLDQMEPTDSSPNHSPAPRPLGSPAPSPAPRPGIQLNLYRPMELIKSL 574
OY 482 SATESLTLLSTGTSTAY--APGLTFNKFYLMKAVEPKSKADENTPPSGEGS----- 531
Db 575 GAAAPHQYFQELPTDLRKPSPOLDEDEPOYKNEVEEKDREHQBMAECSEPEREPL 634
OY 532 -AIAGVADSSATRMQSK-----LVTSLPSMALLTHLKLSTGFFPPVYLEPL----- 579
Db 635 PLEVRKKEEVEDEQVQKQEDHRIEPRTPSPSSSEHSPHHNRHSHMGVPPVQPIQPA 694
OY 580 ---GASPSSTKLOOL-----VEKIDRGAAVAVASTAGAP 612
Db 695 LMRPGSSPGSQSHDLPLTPEQLPPREDFAERFPLNFTTAKMLSEHNSPVASPRAGAL 754
OY 613 TTSAPAP----- 619
Db 755 PPGVPPRRHHNHHMARSPPFNPIKHEMALLPRHSNDNSMENFEVSNTCETMKLKL 814
OY 620 ---SSASGRPOCYCLAVLSCPRALRLHYGONGGERPCKYCGRAFSRGNIRANHVGH 677
Db 815 MKKKKISDPKOCVCDVLSCKSALOMHYRTHTGPERFCRIGRAFTTKGNLKTMAVH 874
OY 678 KTSAPARAQNSCPICOKKFTTAVALTLOOHVHMILG-----GQ 713
Db 875 KIRPPRRNRFQCPYCHKKYSNALVLOQHILRLHTEGPTDLTPQIOAELTRDPPSMNPGH 934
OY 714 IPNGSGALSEGGAAQDENSSEOSTASGSPFPQ-----OSQOFPSEEMSEEE 762
Db 935 FMPRFAAARHFGA-----LPGGPGPPGPNHGAHNGALGSSSQGDMDDNMDCGE 985
OY 763 EEDP-----EEDP-----VTDEDLAGR 781
Db 986 DYDDDDVSSEHLNSNLSOEGRSRSRGGDFKSLLEFQKLRIDATGVVNTNPRRSSASSH 1045
OY 782 GSESGEKA---ISVRDSEEV-----SGAEFEVA-----TSVAAPTYYKEMS----- 822
Db 1046 GHSVGSTAPTSPSVHAASQYIKRSSSPARKSASGALDLTRAAFTSSSSSSSPLPKK 1105
OY 823 -----NEKAPQHTLPPPPP-----PDNLDH----- 843

191 QIHOMQTEOIC---ROVLLIGSLGQTVGAP-----ASPEL----- 224
 450 SPEKIDKREKMKIKKSPRLPGSIREENGVMHVCYCKAEKFRKPDVLRHRIHTHEK 509
 225 ---PGTGAASSTKPL---PLSPKPAQTKTTASSSSSS---SSGAEPKOA 270
 510 FKPCQCFRAFAVASTLTAHKTHTGKAFKCOYCKMSFSTGSLKVIHRLHTGRP--- 565
 271 FENLHPLGSGHPFSGVGVRSH-----KP-----TPAPSPALPGS 306
 566 -PACPH---CDKFKRTSGHKKHTASHKHTELKRMHOKRPKRVKGTNPVPDPLPQ 621
 307 TDOLIASPHLAFPGTGLLAQCSGAAGLEAASAPGLKPKNGSELAGY-GEVYISLEK 365
 622 EPLITDLDLIP-----IPKQFQSYNNNNVNNENDR 655
 366 PGRHKRCRCACAFVGSLSALQILHRSHTGERPKVCNVCNRPFTTGNLKVHFRH---- 420
 656 P---YKCFYCHRAYKKSCHLKQHIRSHTEKPKFCSCGRGFSAGVLKAMHIRTHTGLKS 712
 421 -----REKYPHQMNPVRPHNDYVITSSGLP-----YGMSPRE 456
 713 FKCLICNGAFVTGSGSLRHNGIHNDRPYCMPCQCKTEKTSLNCKNMKTHRYELAQQL 772
 457 KAEBAAGTPGGVERKPLVASTALASATESLTLSTGTAVAPGLPTENKFVLMKAVEP 516
 773 OHQQAASIDSTVDOOSMASTOMOVEIESEDELPTOTAEVYAANP-----EAMLDLEP 824
 517 KS-----KADENTPPSGCSAAGVADSGSATNMQLSKVTLSPSWA 558
 825 QHVNGTEAGLGOOLADQPLEADEDEGVAPODPLRGHVOFEQSPQAQSFEPAGLPQGF 884
 559 LLTNLHKSFGSPFPVYLEPLGASPSERSKIQOLYEKIDRQGVANVAASTGATTSAPA 618
 885 TVTDYVYHQOPFPVYQQLQ-----DSTLESQALSTSFHQOSILOAPSSQGMNVTLLRIQ 939
 619 PSS-----SASGPN-----QCVLCILRVLSCPRALRLHYQGHGERPCKKV 659
 940 ESSQERLDLQAOQSSQFLEDNEDORSRYRCQCNCKGFKKSSHLKQHNHSHTEKRPCKL 999
 660 CGRAFSTRGNLRAHFVGHKTSPPAARAONSCPTQCKKFTNAVTLQOHVRLHMLGQIPNGGS 719
 1000 CGRGFSVSVGLKSH---EKHTHTGVKA--FSCSYCNASFTTNGSLTRHMTATMSMK--PYKCP 1054
 720 ALSEGGCA---AQENSEQGTASGSPFPQSQSPSPREEMSEEEEDD---EEDEEDVTD 774
 1055 FCEGGRTTVVHCKKHKMKRHQTV-----PSAVSANGETEGGDICOMESEEEESD 1101
 775 EDSLAGRSGSESGEKAISVRGDSSEVSGAEDEVATSVAAPTVKEMDSNEKAPQHTLPPR 834
 1102 RNA-----SRKSREPVITTEETA----- 1121
 835 PPPPDMLDHPQMEQGTSD---VSGAMEEAKLGLGISPMALTOEGEGTS--TPLYEE 888
 1122 ---QIAKIRPOESAVSEKVLQVQSAEKD--RISELRLDQALODEPRKANCTCYCPKS 1175
 889 LNLPEAMKMDGSSSGK--ACEVCGOSFPQTALFHKDTHKDKDPLFCVCRQGLD 946
 1176 FKPSDVLVHRVIRHTGKRYKCDGCKSFYKSTLDCVHTKTHGQ--KLFSCHVCSNAPST 1234
 947 RATLKKHMLAHNQVPPF-APV 967
 1235 KGSLLVHMRL-HTGAKRFKCPH 1255

GN ZNF341.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NX NCBI_TaxID=9606;
 [1]
 RP MEDLINE=2163749; PubMed=11780052;
 RX Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 Jones M., Stavrides G., Almeida J.P., Babbsge A.K., Bagguely C.L.,
 Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R.,
 Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
 Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 Lehasalaino M.H., Leyerisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 Marsh V.L., Martin S.L., Mccomachie L.J., McElay K., McMuray A.A.,
 Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showmkeen R., Sims S.,
 Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 Swan R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 Tracey A., Tromans A.C., Vaidin M., Wall M., Wallis J.M.,
 Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 Rogers J.;
 RT "the DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [2]
 RP SEQUENCE OF 227-773 FROM N.A.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
 RA Nimomiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL050349; CAC14087.2; -;
 DR EMBL; AK027550; BAB55193.1; -;
 DR HSSP; P07248; 2ADR.
 DR Genew; HGNC:15992; ZNF341.
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF00096; Zf-C2H2; 13.
 DR SMART; SM00355; Znf_C2H2; 13.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 12.
 KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;
 KW DNA-binding; Repeat.
 FT ZN_FING 42 65 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 227 251 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 269 291 C2H2-TYPE.

816 TVKEM-DSNEKAQHTLP PPPPPDNLDPQMEEGTSVDVGAMEEFAKL-----EGIS 868
::: | : : | | : | : | :
691 GLEELKDTGAGIVPEAVPGKP-----FAEPDAVITVGVGVGA FMTIIVRGCAAGTC 744

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13

PROSITE: PS0157; ZINC_FINGER_C2H2_2; 13.
Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
Nuclear protein; Repeat; Alternative splicing.
DOMAIN 44 125
DOMAIN 310 321
DOMAIN 356 371
ZNF_FING 356 378
ZNF_FING 384 406
ZNF_FING 412 434
ZNF_FING 440 462
DOMAIN 463 484
DOMAIN 485 507
ZNF_FING 513 535
ZNF_FING 541 563
ZNF_FING 569 591
ZNF_FING 597 619
ZNF_FING 625 647
ZNF_FING 653 675
ZNF_FING 681 703
ZNF_FING 709 731
ZNF_FING 731 749
VARSPIC 250 257
CONFLICT 304 305
SEQUENCE 734 AA; 82036 MM; 2BE7D69B18F29437 CRC64;
AL -> RV (IN REF. 1).
EAGGIESP -> MNGPLVYA (IN ISOFORM MZFL/C).
Query Match 5.98; Score 311; DB 1; Length 734;
Best Local Similarity 19.4%; Pred. No. 5.9e-07;
Matches 204; Conservative 74; Mismatches 330; Indels 444; Gaps 39;
11 LGPGCEPERERG-----DASEEH-----POYCA--KC--CAQSPDPEFLAHQ 51
6 LGSPRAPPEDEGPVAVKLEDESEGEALMDPGEAARLRRCFYEATGQALAQ 65
52 NSCTDPVAVIIGQENPSSASAPRPEGHRSQVMDT-----EHSNPD--SGSS 103
66 RELC-----RQMLRPEVRSKQEMLELVLEQFLGALPEIQARVQ 105
104 GPPDPWGERGESESGOFLVATGTRAGG-----GGLIASPKIGAT--PLPP 152
106 G-----QRPSPPEAALVLDGLRRPGRRWTVVOVGQEVLEKPESSFOPLPE 157
153 ESTAPPPPPPPPPPGVSGHNLPLLEELRVLOQROI-----HOMQMTQICROVLL 207
158 TEPPEPPEPKPP-----RTQESPLGLQVKESEVETD----- 192
208 LGLIGQVGAAPASPSBELGTGAASSTRKPLPLSPDKPAQGTKTASSSSSSSGAEP 267
193 -----SDPLESGPLAATQESVPLLP--EAAQRCGTVLQDIPHSKTPGEGP 237
268 KQAFTHLYPLGSGQHPSPVSGVSGSHKTPAPSPAL-----PGSTDQLIASPHLAF 319
238 SWR-----EHPRALMHE--EAGGIF-----SPGFALQIGSISAGPS-----VSPHLHP 280
320 GTTGL--LAAQCLGAARLEMAASPGILKPKNGSGE-----LGYEVVISLEKP 366
281 WDLGAGAGLQGIOSPSR--EGGFHALLPLDLRSQDPDDEPCRGVALLITRMWSP 338
367 GGRH-----KCRFCAKVFSGDSALQIHLRSHTGPRPKCANNCGNFTTNGN 412
339 RGRSRGRPSGTGGVNRGRCRDYCGVFSQSRMLRHOKIHTGERPFVCSSEGRSFRSRS 398
413 LKVHPIRHPKHYVYVQNNPHVPEHLVDVITSSGLPYGMSVPRPEKAEEMAGTPGGVBRK 472
399 LIRHQLIHTHEERFV-----GDCGCG----- 420
473 PLVASTATLSTATSLTLLSTGTAVADGLPTFNKFLMKAVEPKSKADENTPPGSEGA 532
421 -----FVRSARLEHRRVHNGEOP----- 439
533 IAGVADSGSATRMQLSKLVMSLSPWALLTNHLKSTGSPFPVYVLEPIGASPSFTSKLOL 592
440 -FRCAECGGGFRGRSN-----LIQHORIHDGDPGP-----GAKP----- 472

QY 593 VEKIDRGAVAVASTASGAPTSAPAPSSA-----SGPNO--CVICL 633
DB 473 -----PAPGAPDEPPGPPPCSCRSFARRAVILEHQAHTTGDKSPGCEVG 519
QY 634 RVLSCPRALRLHYGQNGGERPFRCKVCGRASTRGNLRHFGHKTSPDARAQNSCPICQ 693
DB 520 ERFRGRSRTLQHRHVSGRFPFACACGGSFRQSRNLTOHRIH-----TGERPFACAE 575
QY 694 KKFTNAVTLQOHVHRNHLGQIRNGSALSSEGAQENSQDSTASGSGFPQSQDS 753
DB 576 KAFQRPRLTQHLRVHTG-----EKP- 596
QY 754 PEELMSEEEDEEEDVDYDSDSLAGSGSGGEKALSVGRDSEVSGAEHEVATVAA 813
DB 597 -----FAC 599
QY 814 PTVKEMDSNEKAPQHTLPPPPPPNDLHPQMEQSTSDVSGAMEEAKLEGISSPMAA 873
DB 600 PEGGQRFSGRLKTRHQRHTGKEP--YHCGEGCLGFTQVS----- 638
QY 874 LTQEGEGSTPLVEELNLEPAKKDPGESSGRKCEVCGSGEPPTOTALIEHOKTHPKDG 933
DB 639 -----RLTEHORIHTGERP--FACPECGSGFRQHANLTQHRRIHTGERP 680
QY 934 LFTGVCRGQFLDRATLKKHMLAHHQVPPA 965
DB 681 -YACPECGKAFQRPRLTQH-LFTHREKPEFA 710
RESULT 12
2133_HUMAN
ID 2133_HUMAN STANDARD; PRT; 654 AA.
AC P52736; O9HA43; O9BUV2;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein 133.
GN ZNF133.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Insulinoma;
RX MEDLINE=95377390; PubMed=7649249;
RA Vissing H., Meyer W.-K., Aagaard L., Tommerup N., Thiesen H.-J.;
RT "Repression of transcriptional activity by heterologous KRAA domains
present in zinc finger proteins.";
RL FEBS Lett. 369:153-157(1995).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levaslatino M.H., Leverisa M.A., Lloyd D., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McIay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Senta H.K., Showkeen R., Sims S.,

RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilmink L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RA "The DNA sequence and comparative analysis of human chromosome 20.";
RLL Nature 414:865-871(2001).
RN (3)
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION AS A
CC REPRESSOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- TISSUE SPECIFICITY: SEEKS UBICUITOUS. SEEN IN THE HEART, BRAIN,
CC PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, AND PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isdb-slb.ch/announce/>
CC or send an email to license@isdb-slb.ch).

CC EMBL: U09366; AAC50260.1; -
DR EMBL: AL049646; CACI5508.1; -
DR EMBL: BC001867; AAH01887.1; -
DR HSSP: P08045; 12NF.
DR TRANSFAC: T04992; -
DR Genev: HGNC:12917; ZNF133.
DR MIM: 604075; -
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF00096; zf-C2H2; 30.
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PR00048; ZINC-FINGER.
DR ProDom: PD000003; ZnF_C2H2; 5.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; ZnF_C2H2; 15.
DR PROSITE: PS50805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 14.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 15.
DR Transcription regulation; zinc-finger; DNA-binding; Metal-binding;
KW Nuclear protein; Repeat; Repressor.
KM
KT
FT DOMAIN 1 72 KRAB.
FT FT 214 631 ZINC_FINGERS.
FT ZN_FING 214 236 C2H2-TYPE.
FT ZN_FING 242 264 C2H2-TYPE.
FT ZN_FING 270 292 C2H2-TYPE.
FT ZN_FING 298 320 C2H2-TYPE.
FT ZN_FING 326 348 C2H2-TYPE.
FT ZN_FING 354 376 C2H2-TYPE.
FT ZN_FING 382 404 C2H2-TYPE.
FT ZN_FING 410 432 C2H2-TYPE.
FT ZN_FING 438 460 C2H2-TYPE.
FT ZN_FING 466 488 C2H2-TYPE.
FT ZN_FING 494 516 C2H2-TYPE.
FT ZN_FING 522 542 C2H2-TYPE.
FT ZN_FING 550 574 C2H2-TYPE.
FT ZN_FING 578 600 C2H2-TYPE.
FT ZN_FING 606 631 C2H2-TYPE (ATYPICAL).
FT CONFLICT 73 73 MISSING (IN REF. 3).
FT CONFLICT 193 193 T->S (IN REF. 2).
FT SEQUENCE 654 AA: 73402 MW: 2D42674529D24228 CRC64;

Query Match 5.8%; Score 306.5; DB 1; Length 654;
Best Local Similarity 20.2%; Pred. No. 8.2e-07;

Matches	172: Conservative	79: Mismatches	277: Indels	323: Gaps	32: 32:
QY	152	PESTPAAPP---PPPPGVGSGHLINPLILELRLVLDQIQIHOMQMTEDICRQ----	204		
Db	68	PATCRADPEPELYLDPEPCPGFSS-----	107		
QY	205	VLLIGSLQGTVGAPASPSLELPGTGAASSTKP-----	252		
Db	108	FTCLCAEGNIQPDGPGDQEKQQAASBGRPMWDQAEGRGEGAMP LF-----	162		
QY	253	ASSSSSSSSSGAPPKQAFPHLYPLIGSQHPFVSQVGRSHKPTAPSPALPSTQOLIA	312		
Db	163	LGAFS-----RPPOR-----	204		
QY	313	SPHLAPGTTGLLAOCLGAARGLAASPDLKRN--GSGE-----	367		
Db	205	RIEVLGFGIVN--CGEC-----	257		
QY	368	GNHK-----	417		
Db	258	ARHQKHAHSEKPIVRCRCGRGFNFKSTLIHERTHSGEKPRYMCSECGRGFSQSNLIHQ	317		
QY	418	HHHREKYPHYQMPHYRPHLDYVITSSGLLYGMYVPREKAEDEACTPGCGVERKPYVAS	477		
Db	318	RTHSGKRP-----	357		
QY	478	TTALSTESTLTLSTGTSTAVAPGLPTFNKPYLMKAVEPKSKADEMTPGSEGSALIGVA	537		
Db	358	DCGLGSDSDSNLIS-----	385		
QY	538	DSGSATRMQLSKLVTSLSMALLTNHLKSTGSFPPEVYLEPLGASPSSTSKLOQLVEKID	597		
Db	386	ECGRCPRRQRTT-----	425		
QY	598	RQGAVAVASTAGAPITSAAPASSASGPNOCVITGLVLSCEPRALRLHNGHGGERPKC	657		
Db	426	---ISHRRTHT-----	468		
QY	658	KVCGRAFTRGNLRAHFVGHKTSPPAARQNSCPICQKFTTNAVTLQOHVMIHGQIP--	715		
Db	469	KCCGRGFSQSNL-----	523		
QY	716	--NGSGALSEGGGAQDENSEDQSTASGPGSPPOSQSPSEEMSEEBEEDDEEEDVT	773		
Db	524	CRECGGFSHQAG-----	552		
QY	774	DEDSLARGSESESGEKALISVRGDEEVSAGAREEVATSVAPPTVTKMDSNEKAPQHTLPP	833		
Db	553	RCQGL--GFNGKS-----	576		
QY	834	PPPPDNLHPQMEQSTDSVSGAMEEBAKLEGISSPMALTOEGEGTSPVLEELNLP	893		
Db	577	-----	579		
QY	894	AMKKDGESSGKACEYCGSFPQTALAEHQTHPKDGRLTCVFCROGLDRAVLTKH	953		
Db	580	-----	624		
QY	954	--MLAHQVP	962		
Db	625	RKTSVHHRLLP	635		

RESULT 13

2219_HUMAN

AC Q9P2Y4; Q9BW28;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Zinc finger protein 219.

GN ZNF219.

OS Homo sapiens (Human).

[illegible]

```

Dh 225 APPPQGPQPPQPPQPPERSVPQPEDEPEDEPEREA-----IPTPARAPADEEPAPPE-- 273
QY 630 VYICLVLSCEPRALRLHYGHGGERPFCKKVCGRAFSTGRLRAHYVGHKTSAPARAQNSC 689
Dh 274 -----FRCCVCGQSFTQSWFLKGHMKHKAS---FDHAC 304
QY 690 PICQKRETNNAVTLQOHVRYMLGGQIPNCGSALSEGGGAQWNSSEGSTAGPSGSPQ- 748
Dh 305 PWCGRCKEPEWFLKNHMKVHASKLGP-----LRAPG-----PASGPAPAPPPD 348
QY 749 -----SOQP-----SPEEEMSEEEED-----EEEDVDYDSDLAGRSGSESG-- 787
Dh 349 LGILATEPLGPAALLIAPAPPPAERREPPSLGYLSLRAGEGRNGGAEEGPGPSGFR 408
QY 788 --EKAISVRC-----DSEYSGAEAEVATVAAPTTVKEMDSNEKAPQHTLPPPP 836
Dh 409 PLSALPARARRHRAEPEEEEEEVEAOEETWARGRSLGSLASL----- 452
QY 837 PPDNDLHPQMEGTSIVYSGAMEEERAKLEGISSPMAALTOEGEGSTPLVEELNPEAMK 896
Dh 453 -----HPRKGE-GPGHSASAGAQR-----STAQEENGL-----LVGGTR 488
QY 897 KDPGESSGRACCEFCGGSFPQTALAEHQKTHPKDPLTCVFCRQGLDRATLKHMLL 956
Dh 489 PEGGRATGKDCFCGGSFSRAHHLKVLRVTGERP-YKCPCHDAGTOSGSLKYH-LQ 546
QY 957 AHH-----QVPPAPHG--PONTATLSLV-----GCSS-STPSGLSPFPPR 995
Dh 547 RHHREORSGAGPPPPPPPSQSGAPQSGAKPSPQAPATWEGASSPPPPSGAGGSR 606
QY 996 KDDPTMP 1002
Dh 607 R-KPASP 612

RESULT 14
Y296_HUMAN
ID_Y296_HUMAN STANDARD; PRT: 1829 AA.
AC 015015;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein KIA0296.
GN KIA0296.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE KRUPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
EMBL; AB002294; BAA20756.1; -.

```

InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 31.
DR PRINTS: P00048; ZINC_FINGER.
DR SMART: SMO0355; Znf_C2H2; 30.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 30.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 29.
KW Hypothetical protein; Transcription regulation; DNA-binding;
Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT DOMAIN 8 1783 ZINC_FINGERS.
FT ZN_FING 8 31 C2H2-TYPE.
FT ZN_FING 48 70 C2H2-TYPE.
FT ZN_FING 75 97 C2H2-TYPE.
FT ZN_FING 239 261 C2H2-TYPE.
FT ZN_FING 266 288 C2H2-TYPE.
FT ZN_FING 294 316 C2H2-TYPE.
FT ZN_FING 374 396 C2H2-TYPE.
FT ZN_FING 401 424 C2H2-TYPE.
FT ZN_FING 465 487 C2H2-TYPE.
FT ZN_FING 492 514 C2H2-TYPE.
FT ZN_FING 575 597 C2H2-TYPE.
FT ZN_FING 617 639 C2H2-TYPE.
FT ZN_FING 644 666 C2H2-TYPE.
FT ZN_FING 821 843 C2H2-TYPE.
FT ZN_FING 848 870 C2H2-TYPE.
FT ZN_FING 881 904 C2H2-TYPE.
FT ZN_FING 958 980 C2H2-TYPE (DEGENERATE).
FT ZN_FING 1052 1074 C2H2-TYPE.
FT ZN_FING 1079 1101 C2H2-TYPE.
FT ZN_FING 1203 1225 C2H2-TYPE.
FT ZN_FING 1230 1252 C2H2-TYPE.
FT ZN_FING 1258 1280 C2H2-TYPE.
FT ZN_FING 1299 1321 C2H2-TYPE.
FT ZN_FING 1326 1348 C2H2-TYPE.
FT ZN_FING 1364 1386 C2H2-TYPE.
FT ZN_FING 1579 1599 C2H2-TYPE.
FT ZN_FING 1585 1607 C2H2-TYPE.
FT ZN_FING 1677 1699 C2H2-TYPE.
FT ZN_FING 1704 1726 C2H2-TYPE.
FT ZN_FING 1732 1754 C2H2-TYPE.
FT ZN_FING 1761 1783 C2H2-TYPE.
SEQUENCE 1829 AA; 200760 MW; B20C462B38684895 CRC64;
Query Match 5.7%; Score 301; DB 1; Length 1829;
Best Local Similarity 20.0%; Pred. No. 3.9e-06;
Matches 168; Conservative 87; Mismatches 291; Indels 296; Gaps 34;
OY 244 KPAQTGKTTASSSSSSSSSGAEPKQAFHLHYHLSQHPSVGVGRSHKPTTPADSPAL 303
DB 796 KPA-TGQPNSSSHSANAVTGQ-----AGAAHTCSDCG----- 827
OY 304 PGSTQOLIASPHLAFGTGTLAAQ-CLGAARGLAAASPGLKFRKNSGELGYEVISS 362
DB 828 -----H-SFPHATGLSHRPC----- 842
OY 363 LEKPGRIHRCFCAYKFGSDALQIHLRSH-----TGERPYKCNVCGNRFRTTGNLKVHF 417
DB 843 --HPPGIYGCSLCPKRFEDSLPALRSHFQNHRRPGFATSQAQPLCCCLCGMFPFGAGYRLH- 899
OY 418 HHRREKPYVQMPHNPVPHLDVYITSSGLPYGMSVPPKAEEDAGTPGGVERKPLVAS 477
DB 900 ---RROAH-----SSSGMTEG-----SEEGEGEVAAAPARSP 931
OY 478 TTAALSTESTLITLSTGTSTAVAPGLPFFNKVLMKAVKPKSADENTPPGSBSAIAGVA 537
DB 932 PLQLSAEELN-----QLQREVALDSAGYGHICGCCGQT---YD 968
OY 538 DSGSATRMOLSKLVLTSLPWSALLTNHLKSTGSPFPYVLEPLGASPSF-----TSKLQOL 592
DB 969 DLGSLER-----HHCQSQSGTTADKAPSPGLGVAGADAMEMVDSVLEDI 1011
OY 593 VKKIDROGAVAVASTAGAPTTSAAPADSSASGPNQCVICLRVLSCPRALRLHYGHGGE 652
DB 1012 VNSVSGEGGDAKQEGAGTPL-----GDSLCT-----GGGE 1042

OY 653 -----RPEKCYGGRAFSTRGNLRAHFVGHGKTSPPAARAQNSCPICOKKFTNAVTLQOH 705
DB 1043 SLLEAQRPRPRCMOCCKTYRHGSLVNHKRIHOTDFL-----CPVCSKCYPLAAVRNH 1097
OY 706 VRMH-----LGGQIPN--GGSALSEG---GGAQENSSEOSTASPGSPPODSQ 751
DB 1098 LRNHPRCKGSEPOVGPPIPEAAGSSELDYGPPIEGGS---NKPOHMAEFGP-----QAEV 1149
OY 752 PPSPEEMSEEBE-DEDEEEDYTDSDSLAGKSF---SGGEFAISVRGSE-----EVSG 802
DB 1150 EKQOEELKVEPLLEVARAVKKEVWEETVKGGEIERLETAEKGCQTEASERPFSEVCG 1209
OY 803 AAEKVNTSV-----AAPTVKEMDSNEKAPOTLPPPPRPNLDHPQMEG. 850
DB 1210 RSYKHAGSLINHROSHOTGHGFCOACSKGFSYLSMLKNH-----RRIHADPRFR 1259
OY 851 TSDVSGAMEEERAKLEGISPMALTOEGESTPLVEE-----LNLPEAMKK 897
DB 1260 CSBGGKAFRLRKQL--ASHQKVHMERGGGGRKATREDRPPRCGGCGRTYRHAGSLNH 1317
OY 898 DPGESSGRKACEVCGSGFPPTQTALEHOKTHPKD-----GLPTCVFCRQGLDRA 948
DB 1318 RSHETGQYSCPTCPKTYSNMALKDHQRLHSENRRRRAGRSRRTAVRCALCGRSFPGRG 1377
OY 949 TLKKHMLAHNOVPPAP-----HGPNINF-----LSLVPGCSSIPS 987
DB 1378 SLERH--LREHETEREPANOGGLDGTASBANLTGSGLFTQLGAEPPVPHLEDGVR 1435
OY 988 PG 989
DB 1436 PG 1437
RESULT 15
ZN84_HUMAN STANDARD; PRT; 738 AA.
AC P51523; O9NNX7; O9UC17; O9UC18;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein 84 (zinc finger protein HPE2).
GN ZNF84.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=89377476; PubMed=2505992;
RA Bellefroid E.J., Lecocq P.J., Benhida A., Poncelet D.A.,
RA "Belayew A., Martial J.A.;
RT "The human genome contains hundreds of genes coding for finger
RT proteins of the Kruppel type.";
RL DNA 8:377-387(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA Grimaldi G.;
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-257 FROM N.A.
RC TISSUE=Teratocarcinoma;
RX MEDLINE=92051312; PubMed=1945843;
RA Rosati M., Marino M., Franze A., Tramontano A., Grimaldi G.;
RT "Members of the zinc finger protein gene family sharing a conserved N-
RT terminal module.";
RL Nucleic Acids Res. 19:5661-5667(1991).
CC -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M27878; AAA79359.1; -
EMBL: X60156; CAB94232.2; -
HSSP: P08047; 1SP2.
GeneW: HGNC:13159; ZNF84.
InterPro: IPR001909; KRAB.
InterPro: IPR000822; Znf_C2H2.
Pfam: PF00096; zf-C2H2; 19.
Pfam: PF01352; KRAB; 1.
PRINTS: PR00046; ZINCFINGER.
ProDom: PD00003; Znf_C2H2; 14.
SMART: SM00349; KRAB; 1.
SMART: SM00355; Znf_C2H2; 19.
PROSITE: PS50805; KRAB; 1.
PROSITE: PS50157; ZINC_FINGER_C2H2_1; 19.
PROSITE: PS00028; ZINC_FINGER_C2H2_2; 19.
Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
Nuclear protein; Repeat.
DOMAIN 8 KRAB.
ZNF_FINGERS.
ZNF_FING 207 733 ZINC_FINGERS.
ZNF_FING 207 229 C2H2-TYPE.
ZNF_FING 235 257 C2H2-TYPE.
ZNF_FING 263 285 C2H2-TYPE.
ZNF_FING 291 313 C2H2-TYPE.
ZNF_FING 319 341 C2H2-TYPE.
ZNF_FING 347 369 C2H2-TYPE.
ZNF_FING 375 397 C2H2-TYPE.
ZNF_FING 403 425 C2H2-TYPE.
ZNF_FING 431 453 C2H2-TYPE.
ZNF_FING 459 481 C2H2-TYPE.
ZNF_FING 487 509 C2H2-TYPE.
ZNF_FING 515 537 C2H2-TYPE.
ZNF_FING 543 565 C2H2-TYPE.
ZNF_FING 571 593 C2H2-TYPE.
ZNF_FING 599 621 C2H2-TYPE.
ZNF_FING 627 649 C2H2-TYPE.
ZNF_FING 655 677 C2H2-TYPE.
ZNF_FING 683 705 C2H2-TYPE.
ZNF_FING 711 733 C2H2-TYPE.
CONFLICT 62 E -> O (IN REF. 2 AND 3).
SEQUENCE 738 AA: 85458 MW: 55268AID98179EED CRC64;

Query Match 5.5%; Score 290; DB 1; Length 738;
Best Local Similarity 21.4%; Pred. No. 4.8e-06;
Matches 135; Conservative 63; Mismatches 222; Indels 210; Gaps 20;

364 EKPGRHKKFCACAVFGSDALQILHRSHTGERPYKWCNGRFTTGNLKVHFRHR-E 422
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
260 EKP---YNCQCGKAFKFSQKSLTSHQRTHTGKPYECGCKAFKFSRSHLISHWRTHTGE 316
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
423 KYPHYQMNPRVPHRLDYVITSSGLPYGMSVPERKAEDEEGTGGGVKRLVASTALS 482
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
317 K-----PYGCGN-----ECG-----RAFS 329
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
483 ATESLTLLSTGTSTAVAPGLPTFNKFLMKAVEPKSKADENTPPSGSALAGVADSGSA 542
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
330 EKSNNL-----NHQRHTGKPEFCEGCKAFKRSKO-----PYLPLELGASPSPT 586
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
543 TRMQLSKLVTSLSPMALLTNHLKSTGSPFP-----PYLPLELGASPSPT 586
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
362 -----LVTHHRTHTGTGKPGSCDKRAFEKSESLIHQRTHTGKPYEC 405
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
587 SKLOQLVEKIDRGCAVAVASTASGAPTTASAPAPSSASGPNOCYICLRVILSCPRALRLHY 646
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
406 SECKRAFR--ERSSLIHQRTHTG-----EKPHGCIQCKGKAFKFSRSHLISHO 450

QY 647 GQHGGERPFCKYCGRAFSTRGNILRAHFVGHKTSPPARAONSCPICOKKFTNAVTLQOHV 706
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 451 MHTTGKEPFLCKGKAFKRSKSL-----VRHQRTHTGKPYECSECKAFKSEKSLSTHQ 506
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 707 RMHLGQIPNGGSLSEGGAAQENS---SEGSTASPGSPQPOQSOQSPSEEMSEEE 763
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 507 RIHTGKE-----PYVSECGKAFKFSRSHLISHQRTHTG-----EKPY 543
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 764 EDDEEEDVDYDEDSLAGRSGSGEKAISVRGDSSEVSGAEAEVATSVAAFTYKEMSN 823
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 544 ESECGKAFKESKSLAHQRTHTGKPYECR--DCEKAFKSKSLINTH----- 589
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 824 EKAPQHTLPPPPPPDNDHPQMEQGSTSVSGAMEEKLEGISPMALTOEGEGTST 883
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 590 -----QRHTGKEPYPYCSLCKRAFEKSESLIHQRTHTG----- 624
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 884 PLVEELNLPAMKKDPGESS-----GRK--ACEVCGQSFPTOTALERHOKTHPKDG 932
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 625 ---KPYECNECKRAFRKRSKSLNHQRTHTGKPYECSECGKAFKFSRSHLPHQRTHTGKE 681
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 933 PLFTCVFCRQGFIDRATLKKMLAHQVP 962
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 682 P-YGCECKRAFSOKSLVNHQRTHTGEP 710
||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 16
DRPL_HUMAN STANDARD; PRT; 1185 AA.
ID DRPL_HUMAN
AC P54259;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Atrophin-1 (Dentatorubral pallidolysian atrophy protein).
GN DRPLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Brain;
RX MEDLINE=95144175; PubMed=7842016;
RA Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,
RA Inoue T., Yamada M.;
RT "Structure and expression of the gene responsible for the triplet
RT repeat disorder, dentatorubral and pallidolysian atrophy (DRPLA).";
RL Nat. Genet. 8:177-182(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96262314; PubMed=8965642;
RA Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,
RA Kidwai A.S., Ashworth R.G., Ross C.A.;
RT "DRPLA gene (atrophin-1) sequence and mRNA expression in human
RT brain.";
RL Brain Res. Mol. Brain Res. 36:219-226(1996).
RN [3]
RP SEQUENCE OF 470-725 FROM N.A.
RC TISSUE=Brain cortex;
RX MEDLINE=93315145; PubMed=8325628;
RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;
RT "Novel triplet repeat containing genes in human brain: cloning,
RT expression, and length polymorphisms.";
RL Genomics 16:572-579(1993).
CC -I- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,
CC Ovary, Testis and Prostate. Lower levels are detected in the
CC Liver, Thymus and Leukocytes.
CC -I- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC
CC (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO
CC ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN
CC EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE
CC DISEASE.
CC -I- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-

CC PALLIDUS AN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE
CC DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE
CC NUCLEUS, RUBRUM, GLOBUS PALLIDUS AND LUY'S BODY. CLINICAL FEATURES
CC ARE MYOCLONUS EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF
CC THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH
CC IN THE FOURTH.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D31840; BAA06626.1; -
CC EMBL; U23851; AAB50276.1; -
CC EMBL; L10377; -; NOT_ANNOTATED_CDS.
CC Genew: HGNC:3033; DRPLA.
CC MIM; 125370; -
CC InterPro; IPR002951; Atrophin.
CC Pfam; PF03154; Atrophin-1; 2.
CC DR PRINTS; PRO1222; ATROPHIN.
CC KW Triplet repeat expansion; Polymorphism.
CC FT DOMAIN 73 82 SER/GLU-RICH (MIXED CHARGE).
CC FT 302 305 POLY-PRO.
CC FT DOMAIN 376 382 POLY-SER.
CC FT DOMAIN 386 397 POLY-SER.
CC FT DOMAIN 442 447 POLY-PRO.
CC FT DOMAIN 479 483 POLY-HIS.
CC FT DOMAIN 484 497 POLY-GLN.
CC FT DOMAIN 504 507 POLY-PRO.
CC FT DOMAIN 564 574 POLY-SER.
CC FT DOMAIN 704 707 POLY-PRO.
CC FT DOMAIN 802 815 ARG/GLU-RICH (MIXED CHARGE).
CC FT DOMAIN 816 827 ARG/GLU-RICH (MIXED CHARGE).
CC FT DOMAIN 925 934 MISSING (IN REF. 2).
CC FT CONFLICT 94 94 Y -> H (IN REF. 2).
CC FT CONFLICT 333 333 M -> I (IN REF. 2).
CC FT CONFLICT 339 339 P -> T (IN REF. 3).
CC FT CONFLICT 541 541 C -> A (IN REF. 2).
CC FT CONFLICT 1028 1028
CC SO SEQUENCE 1185 AA; 124785 MW; 56c306287331c005 CRC64;

Query Match 5.5%; Score 290; DB 1; Length 1185;
Best Local Similarity 21.2%; Pred. No. 7.6e-06;
Matches 234; Conservative 81; Mismatches 399; Indels 392; Gaps 45;

QY 42 SPTPEFLAHONSCCTDPVAVYIGGQENPNSNSASSAPRPEGHSRSQVMDTENSPPDS-100
Db 119 SDPRDI--DODNKRSTSPSI-YSPGSEVNDSDSSGSLGQGPAPRYHPRLPPSPQPDST-175
QY 101 -----GSSGPRDPDWGPERRBEESGQFLYA-126
Db 176 PROPEASFEHPHSVTPPGYAHMERPTSRMFGAPRCAPRPHPOLYRGTGCVGLSGPRMGP-235
QY 127 ATGTAAAGGGG-----LILASPKLGTPLRPESTR-----APP-----160
Db 236 KGGGAASVSGGPRNGKGRHPRTTPIVSASSGASGAPRTKRPPTTPVGGGGLNSAPRPAPFP-295
QY 161 -----PPPP-----PPPGVGS---GHLNPLILEELRLVLOQROIHOMQMTQEI-201
Db 296 HTVPTNLPRLPALPRLNNASAPPGIGAPRLPGLHTSPRY-----333
QY 202 CQOVLLGSLGQRYG-----PAPSGSELPGTGAASSTKRLPLFSPILKPATG-249
Db 334 -----AMGQGGGGLPGRPEKGPETLAP-SPHSLP--PASSAAPAPPMKFPYSSSSSS-381
QY 250 KTTASSSSSSSSSGAE--PRQAEFFHLNPLGDSQHFPVSGVGSRSHKPTPAPSPALPGST-307
Db 382 SAAASSSSSSSSSSASFPFAGDALPSTYHSHFPPTPSLW-----SNQPKYTGQSLP---433
QY 308 DQIASPHLAFPGTGTLLAQCGLAARGLAALAAAPGLLRPKNSGELGVEIVISLEKPG-367

Db	434	SOAWMSQGRPPRRPYGALLA-----NSNAHDGPRPPTGAQ-----	STAHPPY	476
QY	368	GRHKRCFCAKVFEGSDSLQIHLRSHTGERPKVCNWCGRNF--TTRGNILVHFHRRREKY	4253	
Db	477	STHHHHHHQOOOQOOOQOOOHHNNSGRPP-----PGAFHRLREGSSHHNHPYAMSP	530	
QY	426	HWOMNHPV-PEHL---DYVTSSGLRPGYSVPEKAEEEAGT-----	464	
Db	531	LGSLRPYPCGPAHLRPHNSGVSYQAQDPNG---PPVSSSSNSSSTQSGYPCSHPSQ	5875	
QY	465	-FGGVERRPLVASTYLASTESTLLST-----GTSTAARGLPTFK-----FVLAK	512	
Db	588	GROGADYPRPPRPVPTVTSSATLS-TVIAIVASSPAGYTPASBPGRPYGKRARSPGAYKT	646	
QY	513	AVERKKADE-----NTPRGSSEGS---AIGVADSGSATRMQLSKLYLPSMALLTJNH	5633	
Db	647	ALPPGKPKGSPRPFRTGTPRGYKRTSPRAGCTPKRGSPT-----	666	
QY	564	LKSTGSFRRPVY-----LEPLGASPESTKLOOLVEKIDROGAVAAVASTAGARTSA	616	
Db	687	---VGGPRLPPRPAQPSGLRPLRPPRAADASGRPLS--ATQIKQEPAREYETPESPVPARS	7411	
QY	617	PAPSSASGPNOCVILRLVLSCPRALRLHYQHNGGERFKKVCGRAFSTRGRLRHPVG	676	
Db	742	PSPPP-----KVVDP-----SHASQSAEFNK	7633	
QY	677	HKTSPPARAQNSPCICQKFTNAVTLDOHVMHILGQIIPNGSALSSEGAQENSSESQ	736	
Db	764	H---LDRGNNSCARSLYF-----VPLESKLAKKADLVEKYRREA	802	
QY	737	TASGSPFPQPOSQPRSEEMSEEEDEEDEDVTDSDSLAGRS---ESGGEKAISV	7933	
Db	803	-----EGRARBEKEREREREREKERERLEERSVKLQAGSRAVVECPSLGVRH	854	
QY	794	RGDSEVSGAEVEEYATSVAAPTTYKEMDSNKKARQHTLPPRRPPNDLHDROMEGTSD	8533	
Db	855	RPFFE--PGSAVAATVPRLTBDPTALRTLSEYARFNVH---SPGNRMHPYVRLGAVD	9075	
QY	854	VSGAMEDEAKLEGISSPMALLTQEGEGSTPLVEELNLRPAKKDPRGESSGKKACVCGQ	9133	
Db	908	-----PGLGYNVALYSSDP-----	923	
QY	914	SPPTQALAEHQKTHPKDGLFTLCVFCRQGLFDLRATLKKHMLDHNQV---PRAPNG	968	
Db	924	---AREREREREREND-----LRDLRLKPGFEVAKRSELEPLHGVGPGIDPPRRHG	970	
QY	969	PONTATLSLVPGCCSSSTIPSPGLSPFP	994	
Db	971	-----GLALQPG-----PGLIHFPF	985	
RESULT 17				
XFIN_XENLA				
AC	P08045;	STANDARD;	PRT;	1350 AA.
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Zinc finger protein Xfin.			
GN	XFIN.			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;			
OC	Xenopodinae; Xenopus.			
OX	NCBI_TaxID=8335;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Embryo;			
RX	MEDLINE=88082679; PubMed=2826129;			
RA	Ruiz I Altaba A., Perry-O'Keefe H., Melton D.A.;			
RT	"Xfin": an embryonic gene encoding a multifingered protein in			
RT	Xenopus."			

EMBO J. 6:3065-3070(1987).
[2]
CHARACTERIZATION.
MEDLINE=94021366; PubMed=7692399;
Andrezsoi M., de Lucchini S., Costa M., Barsacchi G.;
"RNA binding properties and evolutionary conservation of the Xenopus
multifinger protein Xfin.";
Nucleic Acids Res. 21:4218-4225(1993).
[3]
STRUCTURE BY NMR OF FINGER 31.
MEDLINE=89346749; PubMed=2503871;
Lee M.S., Gipeert G.P., Soman K.V., Case D.A., Wright P.E.;
"Three-dimensional solution structure of a single zinc finger DNA-
binding domain.";
Science 245:635-637(1989).
[4]
STRUCTURE BY NMR OF A FINGER.
MEDLINE=89378224; PubMed=2506074;
Lee M.S., Cavanagh J., Wright P.E.;
"Complete assignment of the 1H NMR spectrum of a synthetic zinc
finger from Xfin. Sequential resonance assignments and secondary
structure.";
FEBS Lett. 254:159-164(1989).
-1- FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSCRIPTIONAL
REGULATION PROCESSES.
-1- SUBCELLULAR LOCATION: CYTOPLASMIC.
-1- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH
AS NEURAL RETINA CONES.
-1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND
EMBRYOGENESIS.
-1- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.
-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
-1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: X06021; CAA29425.1; -
PIR: S00647; S00647.
PDB: 1ZNF; 15-OCT-91.
InterPro: IPR001909; KRAB.
InterPro: IPR000822; ZnF_C2H2.
Pfam: PF01352; KRAB; 1.
PRINTS: PR00048; ZINCINGER.
PRODOM: PP000003; ZnF_C2H2; 20.
SMART: SM00349; KRAB; 1.
SMART: SM00355; ZnF_C2H2; 35.
PROSITE: PS50805; KRAB; 1.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 35.
PROSITE: PS50157; ZINC_FINGER_C2H2_2; 37.
Zinc-finger: Metal-binding; RNA-binding; Repeat; 3D-structure;
Phosphorylation.
DOMAIN 1 58 KRAB.
DOMAIN 108 298 ZINC-FINGERS I.
DOMAIN 326 488 ZINC-FINGERS II.
DOMAIN 503 721 ZINC-FINGERS III.
DOMAIN 750 940 ZINC-FINGERS IV.
DOMAIN 988 1066 ZINC-FINGERS V.
DOMAIN 1136 1298 ZINC-FINGERS VI.
ZNF_FING 108 130 C2H2-TYPE.
ZNF_FING 136 158 C2H2-TYPE.
ZNF_FING 164 186 C2H2-TYPE.
ZNF_FING 192 214 C2H2-TYPE.
ZNF_FING 220 242 C2H2-TYPE.
ZNF_FING 248 270 C2H2-TYPE.

FT	ZNF_FING	276	298	C2H2-TYPE.
FT	ZNF_FING	326	348	C2H2-TYPE.
FT	ZNF_FING	354	376	C2H2-TYPE.
FT	ZNF_FING	382	404	C2H2-TYPE.
FT	ZNF_FING	410	432	C2H2-TYPE.
FT	ZNF_FING	438	460	C2H2-TYPE.
FT	ZNF_FING	466	488	C2H2-TYPE.
FT	ZNF_FING	503	525	C2H2-TYPE.
FT	ZNF_FING	531	553	C2H2-TYPE.
FT	ZNF_FING	559	581	C2H2-TYPE.
FT	ZNF_FING	587	609	C2H2-TYPE.
FT	ZNF_FING	615	637	C2H2-TYPE.
FT	ZNF_FING	643	665	C2H2-TYPE.
FT	ZNF_FING	671	693	C2H2-TYPE.
FT	ZNF_FING	699	721	C2H2-TYPE.
FT	ZNF_FING	750	772	C2H2-TYPE.
FT	ZNF_FING	778	800	C2H2-TYPE.
FT	ZNF_FING	806	828	C2H2-TYPE.
FT	ZNF_FING	834	856	C2H2-TYPE.
FT	ZNF_FING	862	884	C2H2-TYPE.
FT	ZNF_FING	890	912	C2H2-TYPE.
FT	ZNF_FING	918	940	C2H2-TYPE.
FT	ZNF_FING	988	1010	C2H2-TYPE.
FT	ZNF_FING	1016	1038	C2H2-TYPE.
FT	ZNF_FING	1044	1066	C2H2-TYPE.
FT	ZNF_FING	1136	1158	C2H2-TYPE.
FT	ZNF_FING	1164	1186	C2H2-TYPE.
FT	ZNF_FING	1192	1214	C2H2-TYPE.
FT	ZNF_FING	1220	1242	C2H2-TYPE.
FT	ZNF_FING	1248	1270	C2H2-TYPE.
FT	ZNF_FING	1276	1298	C2H2-TYPE.
FT	STRAND	1045	1045	
FT	STRAND	1052	1052	
FT	HELIX	1056	1065	
FT	TURN	1066	1066	
SEQUENCE	1350 AA; 155804 MW; 27F10AB0851E0AD8 CRC64;			
Query Match	5.5%; Score 290; DB 1; Length 1350;			
Best Local Similarity	20.9%; Pred. No. 8.6e-06;			
Matches 126; Conservative 61; Mismatches 248; Indels 168; Gaps 18;				
QY	364	EKPGRHRCFKFAKVGSDSALQHLMSHTEGRRPKVCNCGRRFTTRGILKVFHRRREK	423	
DB	407	EKP---FKCSHCDKRTKTERALKHQRTHGEKPKSCDCKEFTORSNLIHQRIHTE	463	
QY	424	YPHVQMPHPVPEHLDYVITSSGLPYGMSVPEKAEBAAGTGGVERRPDYASTALSA	483	
DB	464	RPY-----	466	
QY	484	TESTLLSTGTSTAVAPGLPTFNKFLMKAVEPKSKADENTPPGSGSALAGVADSGSA	542	
DB	467	--KCTLCD-----RTFIONSDLVHKQKHANLP-----LSDPTAN	500	
QY	543	TRMQLSKIVTSLPSMALLTTHLK-STGSEFPFVYLEPIGASPSKTLQVLVEKIDRGA	601	
DB	501	SPHKSCDLPFSHWTFPMKSHKHSEKRP-----QCAECK--GFTOKSDLVKH	549	
QY	602	VAVASTASGAPTTASAPSSASGPNOCVILRYLSCPRALRYHGOHSGGERPFCKKVCG	661	
DB	550	IRV-----HTGEKPFCKLCKKSSQNSDLHKHMRINHGEKFPFCYTC	593	
QY	662	RAFSTRGNLRAPHVGHKTSFAPRAQNSCPICQCKFTNAVTLQOHVHMLGGQIPNGSAL	721	
DB	594	KSFTERSALIKH--HRTHTGER-PHKSCVQCKGFIQKSLTKHSRTHG-----	639	
QY	722	SEGGAAQNSSEQSTASGSGSPFQPOSQSPREPMSEEEDEDEEDVDYDDEDSLAGR	781	
DB	640	-----EKPYPCQCK--SFQNSDLVKNRHTHGEKPNHCTEKNKRFTEGSSLVKH	689	
QY	782	GSESGEKALISVGDSEEVGAEAEVATVAPTTVKEMDSNEKAPQHTLPPPPPP---	838	
DB	690	RTHSGEK-----PYRCPCQCKFTFQSSDLVKNLHVHNGENPPAATAFH	733	

OY 839 -----DNLDPQPMEOGTSVDSGAMEE- AKLEGISPMALTOEGECTSPLYVELNL 891
DB 734 EILIRRENTRESEPDYPCTECKVFHQRPALLKHLRTHKTEKRYPCNCDKXSEFQTSDL 793
OY 892 PEAMKKDPRESSGRKACEVGGSGFPTOTALAEHOKTHPKDPLFCVFCROGLDNLATLK 951
DB 794 VKHLTHNGERYH--CPECNKGFIQNSDLVKHORTHNGERP--YTCOSQCKGFIQRSALT 850
OY 952 KHM 954
DB 851 KHM 853
RESULT 18
HBL1 CAEEL
AC 09XYD3: 019389; STANDARD: PRT: 982 AA.
DB 16-OCT-2001 (Rel. 40, Created)
DB 16-OCT-2001 (Rel. 40, Last sequence update)
DB 15-JUN-2002 (Rel. 41, Last annotation update)
DB Hunchback-like protein.
GN HBL-1 OR FL3D11.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN
RP SEQUENCE FROM N.A.
RP STRAIN=Bristol N2.
RX MEDLINE=99117349; PubMed=917360;
RA Fay D.S., Stanley H.M., Han M., Wood W.B.;
RT "A Caenorhabditis elegans homologue of hunchback is required for late
RN stages of development but not early embryonic patterning.";
RN Dev. Biol. 205:240-253 (1999).
RP
RP SEQUENCE FROM N.A.
RP STRAIN=Bristol N2.
RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RL
CC -1- FUNCTION: REQUIRED FOR LATE STAGES OF DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN ECTODERMAL CELLS DURING
CC EMBRYONIC AND LARVAL DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF09737; AAD16170.1; .
DR EMBL; U40939; AAAB1701.3; .
DR Wormpep; FL3D11.2; CE23646.
DR InterPro; IPR000822; Znf.C2H2.
DR Pfam; PF00096; zf-C2H2; 9.
DR SMART; SM00355; Znf.C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KW Developmental protein; zinc-finger; Metal-binding; DNA-binding;
KW Repeat; Nuclear protein.
FT DOMAIN 336 981 ZINC FINGERS.
FT ZN_FING 336 358 C2H2-TYPE.
FT ZN_FING 361 384 C2H2-TYPE.
FT ZN_FING 538 560 C2H2-TYPE.
FT ZN_FING 567 589 C2H2-TYPE.
FT ZN_FING 595 617 C2H2-TYPE.
FT ZN_FING 623 647 C2H2-TYPE.
FT ZN_FING 734 756 C2H2-TYPE.
FT ZN_FING 929 951 C2H2-TYPE.

FT ZN_FING 957 981 C2H2-TYPE.
SQ SEQUENCE 982 AA; 106951 MW; 49812554CF3CE30F CRC64;
Query Match 5.48; Score 287.5; DB 1; Length 982;
Best Local Similarity 18.88; Pred. No. 8.1e-06;
Matches 211; Conservative 113; Mismatches 408; Indels 393; Gaps 38;
OY 2 ADETGSSRLGCGPGFAERGDASEEHNHPQVCAKCAQFSDPTFLAHONSCCTDPVM 61
DB 85 AQPGEKIHDPDGYVSPKEDGRKSS-----HTNS----- 114
OY 62 VIIGGEMPNSSASSAPRPEGHSRQVMDTEHNSNPPDSGSGPPDPWGPERGERSSG 121
DB 115 YDVASQSPENDGAQSDSTSEHDIVBCEMTETEMDTDEKDTFKPEDQATPKLEEGSDSK 174
OY 122 QPLVAATGTAGAGGGGILLASPKIGATPLPRESTAP-----PPPPPPPPPPVG 171
DB 175 PESTVEGTS-----SNVQVTSFVOMPMQIPVIPSFLKNSLPAPIPPTTQSAN 225
OY 172 SGHLNPLILEELRVLQORQIHOQMTPQICRQVLLGLGTVGAPASPSLPGTGAAS 231
DB 226 VERNSNPSTIEALL-----TLSQOQFAEVFAEAAIKRSSSESIGQR-----GTSAPL 276
OY 232 STKPLPLFSPK-----PAQTGKTATSSSSSSSSGAEPKQAFPHLYHPLGSOHFS 285
DB 277 NIEPKEMSSANNNNEAPASTVASCTPTTTSASFRCRP----- 318
OY 286 VGVGVSHKRTTPASPLPGSTDQILASPHLAFPGTGTLLAQCIGAARGLAASGGL 345
DB 319 --GLG--PVALP--PQNGOTPMV----- 337
OY 346 KPRNGGELGYEYISLEKGRHKRCFAKVFSGDSALQILRHNTGTERPKCNVCGN 405
DB 338 -----CPICGMCSSKRFNPNHMH--GD--HQCSCDY 368
OY 406 RETTGNLKVHFRHREKYPHVQNPHPVPEHLDVYITSSGLPYGMSVPEKAEERAGTP 465
DB 369 TSTRERLKKHM--RES-----HTVEQLRAGFES-----EPAKESASP 406
OY 466 GGVGEKPLVASTALATESLTJLSTGSTVAAGLPFNKFLMKAVPKKADENTP 525
DB 407 -----KNLSLSKDSAT--SPINELFNSTWASI----- 434
OY 526 PGSEGAIAGVADSGASATRMOLSKLVTSLSWALTNHLSKSGSFPPVLEPLGASPE 585
DB 435 LDSTNNAV-----SSTTEPSAL-----SALTDMST-----PILSTLHNSFG 477
OY 586 TSKLQQLVEKIDR-----QCAVAVAS--TAGAPTTAPAPSSASGP----- 626
DB 478 VSLDQIKAISENSPFPEGGINLASLGVSNALIKGDPPEKQSGECRRSSGKIKI 537
OY 627 ---NQC-----VICLAVLSCPRA-----LKLHGQHGGERFKC 657
DB 538 FKCKQGHOSLKDQWANAHTTHIPAEKQLNQCQHNCEVTEYKHHLEYHVRNHLSGSKPQC 597
OY 658 KYCGRAFSTRGNLRAHVGHKTSAPARAONSCPICOK----- 694
DB 598 KCAAYNCVKNKSMNLNKKSHNHYQRCMD--CTYATKYHSLKLHLKTYNHRVPECIEM 656
OY 695 -----KFTNAVTLQ 703
DB 657 SGGDSPPFTSDATITPSPILKKOEIKETVEPVTSIAQFPFPMGMGNHGLNFAHMLN 716
OY 704 QVRYMHLGG-----QIRNGSGALSDEGGAQOE-----NSSE--OSTASGPGSFPQ 746
DB 717 KHLVDYGLMGLRNSVYMSPLKSCACDFVASSADEKMRHSMHLLNSVVPSTIASLYSNL 776
OY 747 PQSQQSPSEEMSEEEDEEDVDVTDSDSLAGSGSEGEKAIIVRGDSEEVSGAE-- 804
DB 777 PFSHYAPPDNDNLESMDCKIDDDNITTESHCYEBMDGGSASVPTSSQSLSSGDEET 836
OY 805 -----EEVATSVAAPTVKKEMDS--NEKAPQHTLPPPPPPDNLDPQPMEOGTSVDS 855

837 KKCKSLSEIOISARANGNNSPMSNDNAMEKDESDADDAPHSPSDTTVPSPPLHSSSIVA 896
856 GAMEEAKLEGTSSPAALAQEGGTSTPIVEELNPEAMKRPSPSSSGKACVCGQSF 915
897 PIPITPPQNEFLQSLIAQASLUG-----PLT--ANRPSAF-----YCHCKIPF 938
916 POTALLEHOKTHPKDGLFTCVCRQGFILDRATLKKHMLLAHQ 960
939 DTQOVLDSHMRFTPCNP-FMCSDCQYQAFNELSFALHMYQARHQ 982

ESQUT 19
EPI_MOUSE
D ZEP1_MOUSE STANDARD; PRT; 2688 AA.
003172:
01-OCT-1996 (rel. 34, Created)
01-OCT-1996 (rel. 34, Last sequence update)
15-JUN-2002 (rel. 41, Last annotation update)
Zinc finger protein 40 (transcription factor alphaA-CRYBP1) (Alpha A-crystallin-binding protein 1) (Alpha A-CRYBP1).
HIVEP1 OR ZNF40 OR CRYABP1.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
SPRAIN-DBA/21:
MEDLINE=95138112; PubMed=7836383; Donovan D.M., Platiorsky J., Brady J.P., Kantorow M., Sax C.M., "Murine transcription factor alpha A-crystallin binding protein 1. Complete sequence, gene structure, expression, and functional inhibition via antisense RNA." J. Biol. Chem. 270:1221-1229(1995).
[2]
SEQUENCE OF 2024-2688 FROM N.A.
TISSUE=Lens epithelium;
MEDLINE=90287161; PubMed=1694016; Nakamura T., Donovan D.M., Hamada K., Sax C.M., Norman B., Flanagan J.R., Ozato K., Westphal H., Platiorsky J., "Regulation of the mouse alpha A-crystallin gene: isolation of a cDNA encoding a protein that binds to a cis sequence motif shared with the major histocompatibility complex class I gene and other genes." Mol. Cell. Biol. 10:3700-3708(1990).
-1- FUNCTION: TRANSCRIPTION FACTOR WHICH BINDS SPECIFICALLY TO THE PALINROMIC SEQUENCE 5'-GGGAATGCC-3' IN THE ALPHA-A CRYSTALLIN PROMOTER.
-1- SUBCELLULAR LOCATION: Nuclear.
-1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH ZINC-FINGER IN-BETWEEN.
-1- SIMILARITY: STRONG, TO HIVEP2.
-1- SIMILARITY: STRONG, TO HIVEP2.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
EMBL; L36829; AAA98810.1; JOINED.
EMBL; L36825; AAA98810.1; JOINED.
EMBL; L36826; AAA98810.1; JOINED.
EMBL; L36827; AAA98810.1; JOINED.
EMBL; L36828; AAA98810.1; JOINED.
EMBL; X68946; CAA48762.1; JOINED.
HSSD; P15822; 3ZNF.
TRANSFAC; T00007; -.
MGD; MGI:96100; HIVEP1.
InterPro; IPR000822; Znf_C2H2.
Pfam; PF00096; zf-C2H2; 5.
PRINTS; PR00048; ZINCFINGER.

DR SMART; SM00355; ZNF_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 165 170 POLY-SER.
FT DOMAIN 407 459 ZINC_FINGERS.
FT ZN_FING 407 429 C2H2-TYPE.
FT ZN_FING 435 459 C2H2-TYPE.
FT DOMAIN 799 802 POLY-SER.
FT DOMAIN 851 854 POLY-PRO.
FT ZN_FING 953 976 C2HC-TYPE (POTENTIAL).
FT DOMAIN 1482 1486 POLY-SER.
FT DOMAIN 2074 2126 ZINC_FINGERS.
FT ZN_FING 2074 2096 C2H2-TYPE.
FT ZN_FING 2102 2126 C2H2-TYPE.
FT DOMAIN 2179 2182 POLY-ASP.
SQ SEQUENCE 2688 AA; 288341 MW; 5EAD46C3A7008BE6 CRC64;
Query Match 5.3%; Score 283.5; DA 1; Length 2688;
Best local similarity 20.1%; Pred. No. 3.2e-05;
Matches 224; Conservative 130; Mismatches 409; Indels 353; Gaps 48;
QY 3 QETGSSSRLLGPGCGEPAERG-----DASEHHDPQVCAKCAQFSDPTFFLAHONSCTD 57
DB 107 QFTKONGETPPGMTAESSESGDLVSPKRTTSPHRSSELPTR----- 155
QY 58 PPVWYIIGQENPNSNSASAPREGHRSQVMDTEHSPNPPSGSSGPDPTWGPERGE 117
DB 156 -----LSGLDQORSSSS-----SKARTDSECSPP--CSTTPSYT----- 192
QY 118 ESSGQFLVA-----TGAAGGGGGLLASPKL-----GATPLPESTPAP----- 158
DB 193 STAFVILKAMEPELSTLSQKSSCAIKTEKLRPKVTGRSPKLNKNSLIDAPNATSPDLV 252
QY 159 ---PPPPPPPPPPVGGSGHLNPLILEELRYLQORQHOMQTEQICROYLLGSLQIV 215
DB 253 VESPCPCPTSYVHVAVASTOKSEQVAAOCVSHLYSSODHLVPLKLSQONOQ--LPGLHGLT- 309
QY 216 GAPAPSELPFGGAASSTKRLPLFSPKIPKQOTGTTASSSSSSSSSGAEPKQAFPHLY 275
DB 310 -----GSLTNLHTLESTK--LEPYNTAVTVSTGLT-----SPSTQVTPPHQ----- 351
QY 276 HPLGSOHPSPVSGVGGRSHKPTPAPSPALPGSTDLIASPHLAFPTTGLLAACLGARG 335
DB 352 -QMDVSPPLSVSPASSTQSP--PGR-----IYSAHVASV-----VSQS 387
QY 336 LEAASPGLLKPRKNSGELGYEYISLEKPGGRHKRCAPKAVVFGSDALQIHRSHTGE 395
DB 388 VEOMCSLLLRDQK-----PKKQKGYICEVCNCAKAPSVLKHIRSHTGE 432
QY 396 RPYKCNVCGNRPFTTRGNLKVHFRHREKYPH-VQMNPPHPLDLVITSSGLHPYGSVP 454
DB 433 RPYCVTCGFSEFTKSNL-----YKHKSHANTTKGLVLOPE-AGGLFLSQECPKALSVH 487
QY 455 ---PKAEAEAGTPGGGVERKPL-----VASTALSAATESLTLSTGTSTAVAPGLPTF 505
DB 488 SDIEGSGEDDESLADGRQNNPCVKDLQPVQTMKTVSNPESLPLIPNSDHYVRGF--- 544
QY 506 NKFVLMKAVEPKSKADENTPPGSEGSALAGVADSSAIRMQSLKLVTSIPSMALLTNHLK 565
DB 545 -----SSQDRPSSQOAPTELPK----- 561
QY 566 STGSPFPVYLEPLGASPSFTSKLQOLVEKIDRGAVAVASVAGAPTSAPAPSSSASG 625
DB 562 -----VYHPVSNPPLKTDCLQ-----VAN-----PNELPSPQS----- 591
QY 626 PNOCVTCLRVLSCPRALRL-----HYGOHGG-ERPFPKCVKGRAPSTRGNLRAHFVGHKT 679
DB 592 -----PRDLHVASILSHSASVSSLEMDSCHOKGDVIGSEKPDH----- 632
QY 680 SPAAQAQNSCPTQCKFTTNAVTTLQOHVBMILGGQIIPNGSGALSSEGGAQOENSSEOSTAS 739

755 SSLKKHKKIHRREKPPKCECGKAFIWSSTLTTRHKRIHTGE--KPYKCECGKAFSSRSST 812
589 LQOLVEKIDROGAVAVASTASGAPTTAPAPSSASGPNOCVICLRVSCPRALRLHYGO 648
813 LTK--HKTIHNG-----EKPYKKECGKAKRHSAALKHKII 847
649 HGGEPFECCKVCGRAFSTGNGNLRAHVGNKTSPPARAONSCPTCOKFTMAVTLQOHVEM 708
848 HAGEKLYKCECGKAFNOSNLTTHKIIHREKPKSSEE--CDKAFIWSSTLTREKRI 903
709 HLGQIPNGSALSBSGGAAOENSSSOSTASGP-----GSFPOQOQSPREEMSE 761
904 H-----TREKPYKCECGKAFQSPHLLTHKRMHTGK 936
762 EEDDEEEDVYDEDSLARGSESGEKAISVRGDSGAEFEVATVAPTVKEND 821
937 PYKCECGKAFSSSTLTTHKIIHTEGKPYKC-----EECGAFKRSSTLTTE-- 983
822 SNEKAPQHTLPPPPPPDLHPQPMEOGTSDVSGAMEEAKLEGISSPMALTOEGEGT 881
984 -----HKTIH-----TGEKPYKCECGKAFSSSTLTTRHTMHTGE 1019
882 STPLVEE-----LNLPEAMKKDDGESSGRK--ACEVCGSPPTQALAHNOKTHPRDGPLF 935
1020 KPYKCECGKAFNRSSSKLTTHKTIHTEGKPYKCECGKAFISSSTLNGHKRIHTEKRP-Y 1078
936 TCVFCROGFLDRATLTKHMLLAHNQVP 962
1079 KCECGKAFSSSTLTTRHKRIHTGKRP 1105

RESULT 22
2169_HUMAN
ID AC 014929;
15-DEC-1998 (Rel. 37, Created)
15-DEC-1998 (Rel. 37, Last sequence update)
15-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 169 (Fragment).
ZNF169.
Homo sapiens (Human).
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI:taxid=9606;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=97225201; PubMed=9071574;
Chidambaram A., Gallant M., Gerrard B., Stewart C., Goldstein A.,
Chumakov I., Bale A.E., Dean M.;
Characterization of a YAC contig containing the NBCCS locus and a
novel Kruppel-type zinc finger sequence on chromosome segment
9q22.3.*;
Genes Chromosomes Cancer 18:212-218(1997).
-1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
-1- SUBCELLULAR LOCATION: Nuclear (Potentially).
-1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN KIDNEY, IS WEAKLY
EXPRESSED IN HEART, LIVER, SPLEEN, AND SMALL INTESTINE, AND IS NOT
EXPRESSED IN ADULT BRAIN OR SPINAL CORD.
-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

CC EMBL: U28251; AAA70188.1; .
DR HSSP: P25490; IUBD.
DR Genev: HGNC:12957; ZNF169.
MIM: 603404; .

DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 11.
DR SMART: SM00355; ZNF_C2H2; 10.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 12.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 1 144 473 ZINC_FINGERS.
FT ZN_FING 144 166 C2H2-TYPE.
FT ZN_FING 172 194 C2H2-TYPE.
FT ZN_FING 200 222 C2H2-TYPE.
FT ZN_FING 228 250 C2H2-TYPE.
FT ZN_FING 311 333 C2H2-TYPE.
FT ZN_FING 339 361 C2H2-TYPE.
FT ZN_FING 367 389 C2H2-TYPE.
FT ZN_FING 395 417 C2H2-TYPE.
FT ZN_FING 449 473 C2H2-TYPE.
FT ZN_FING 479 504 C2H2-TYPE.
SQ SEQUENCE 512 AA; 57651 MW; 5B7BDF763C63E9A2 CRC64;

Query Match 5.3%; Score 278.5; DB 1; Length 512;
Best Local Similarity 20.6%; Pred. No. 1e-05;
Matches 159; Conservative 66; Mismatches 252; Indels 295; Gaps 28;

QY 221 PSELPTGGAASSTKPLPLPLESPIKPAOTGKTTASSSSSSSSGAPPRQAFFHLHYPLGS 280
DB 2 PAKFPPPGAFSSQDLROYALSGHPQT-----PPSSAGGD-----FQLEAPRCS 47
QY 281 QHPFSVGVGVRSHKPPAPSPAL---PGSTDQLASPHIAPP-----STTGILAAQ 328
DB 48 SEKGSSE-----ETEGPDSSLKKRPSRISRTFFSPHQGPVHEVEGREGTDRLAQ 100
QY 329 --CLGAA---RGLEAASDGLKPKNGSGLGYGEYISSLEKREGKRCRPAKAYGSD 382
DB 101 RMSLGSDPTMLKGADTSESGAVIR--GNRYRLGLSKR--SLFPHOKHNVCPCGGRFCOR 156
QY 383 SALQHLRSHTGERPYKCVNCGNRFTRGMLKVHFHRRREKVRHYQMHPVRHNDYVI 442
DB 157 SDLIHQRTHTGKRPKLCRECGRRFSQKAPSPHQKHSGKP-----YVC 202
QY 443 TSSG--LPRGMSVPRPKAEEDAGTP-----GGGVKRPVASTATLSATESLTLLSTGT 494
DB 203 RECGRHFRYTSISLTINHKRIHSGRRPFVYQCGRGFRQK-----IAL-----LLHQRT 249
QY 495 STAVAPGLPTFNKKFVLMKAVEPKSKADENTPRGSEGSALGAVADSGSATRMOLSKLVYSL 554
DB 250 HLEEK-----FVCLSVGEALOKA----- 268
QY 555 PSMALLTNHLKSTGSPRPVYVLEPLGASPSGTSKLOLVEKIDROGAVAVASTASGAPTT 614
DB 269 ---SLIQHOSSHTGEPF-----LSIEGRTFRQOQSL--LSHQ----- 302
QY 615 SAPAPSSASGPNOCVICLRVSCPRALRLHYGONGGERPPKVCGRAFSTGNGNLRAHF 674
DB 303 ---VTHSGEKRYVCAEGCHSFQKVTLLNHQRTHTGKRPKLCQCGNGFSQKYL---- 354
QY 675 VGHKTSPPARAONSCPTCOKFTMAVTLQOHVNRHLGCGQIPNGSALSBSGGAAOENSS 734
DB 355 IGHQRTHTGKRPYLCRDCRGFGQNGTLIRHQRTHTG----- 391
QY 735 QSTASGSPFPQOQSPREEMSEDEEEDVYDEDSLARGSESGEKAISVR 794
DB 392 -----EAPYLCPKGRAGFKSLTLR----- 412
QY 795 GDSEVSGAEEVATVAPTVKENDSNEKARQNTLPPPPPPDLHPQPMEOGTSDV 854
DB 413 -----HQRTHS----- 418
QY 855 SGAMEEAKLEGISSPMALTOEGEGTSTPLVEELNLPEAMKKDPGESSGRK--ACEVCG 912
DB 419 ---EELLYVDRVCG-----QGLGOKSHLI-----SDQRTHSGERPCICDECG 457

913 OSPTOTALAEHOKTHPKDPLFTVCFCROGFLDRATLKHMLL--AHHQVP 962
 458 RGFEFKSLIRKHTHSEKP-YVCEGCGFSOKSHLRHRTKSGHLLP 508
 RESULT 23
 151.CHICK
 2151.CHICK STANDARD; PRT; 706 AA.
 090625;
 30-MAY-2000 (Rel. 39, Created)
 30-MAY-2000 (Rel. 39, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Zinc finger protein 151 (Zinc finger protein Z13) (Fragment).
 Gallus gallus (Chicken).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 Gallus.
 NCBI_TaxID=9031;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=9603919; PubMed=7575457;
 Schultz T.C., Hopwood B., Rathjen P.D., Wells J.R.E.;
 "An unusual arrangement of 13 zinc fingers in the vertebrate gene
 Z13.";
 Biochem. J. 311:219-224(1995).
 - SUBCELLULAR LOCATION: Nuclear (Potential).
 - SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 FINGER PROTEINS.
 - SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; U14555; AAN21556.1;
 HSSP; P15822; IBD0.
 InterPro; IPR000210; BTB_POZ.
 InterPro; IPR000822; Znf_C2H2.
 Pfam; PF00096; zf_C2H2; 13.
 ProDom; PD000003; Znf_C2H2; 2.
 SMART; SM00355; Znf_C2H2; 13.
 PROSITE; PS00097; BTB; PARTIAL.
 PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
 PROSITE; PS0157; ZINC_FINGER_C2H2_2; 13.
 Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 Nuclear protein; Repeat.
 NON_TER 1
 DOMAIN <1 12 BTB.
 DOMAIN 205 640 ZINC_FINGERS.
 ZN_FING 205 227 C2H2-TYPE.
 ZN_FING 233 255 C2H2-TYPE.
 ZN_FING 261 283 C2H2-TYPE.
 ZN_FING 289 311 C2H2-TYPE.
 ZN_FING 317 339 C2H2-TYPE.
 ZN_FING 345 367 C2H2-TYPE.
 ZN_FING 373 395 C2H2-TYPE.
 ZN_FING 401 423 C2H2-TYPE.
 ZN_FING 427 450 C2H2-TYPE.
 ZN_FING 457 479 C2H2-TYPE.
 ZN_FING 485 507 C2H2-TYPE.
 ZN_FING 513 536 C2H2-TYPE.
 ZN_FING 618 640 C2H2-TYPE.
 SEQUENCE 706 AA; 76820 MW; 4BDAC68808F136 CRC64;
 Query Match 5.28; Score 274; DB 1; Length 706;
 Best Local Similarity 20.58; Pred. No. 2.2e-05;
 Matches 175; Conservative 82; Mismatches 277; Indels 320; Gaps 36;
 181 LEEINVLAQRIHQMG-MTRDQICROVLLIGSLGQ--TVGAPAPSPSELPGTGAASSTKPLL 237

DB 1 LKSLTLVAESVSESELTELGEKVIIVEGKTARVTOGSDSKROYV----- 48
 238 PLSPPIPAQOTKTTAS--SSSSSSSGAEP--KQAFHLHYPLGSGQPFPSVGVGRSHK 294
 49 -----PNOGKEEPAYATAOPKPEAPDPAKEG-----PAEGDP--GGVDNAE 92
 295 PTPAP-SPALPGSTPQLALASPHLAPFGTGLLAQCLGAARGL-----E 337
 93 ASPAAVSPSRPQPAEVEGN---SSPGEKSDAPST--EARGMELEGKEDEEAWVEDE 147
 338 AAASPLGLKPK-----NGSGELG-----YGEVTSLEKPGGRHKC 372
 148 EAKTPKAAQPKPSEKEMEDNESGSTDSGENSEGSTRLRREGTYSDFRESKAYAAVTHKC 207
 373 RFCAKAVGSDSALQIHLNSHTGERPKVCNVCGRFTTGNLKVPHRHKREKPHVQMNH 432
 208 EDCGKEFTHTGTFKRIHRIHTGEKPKSCRCNKAFSDPAACKAHEKTH----- 255
 433 PVPEHLVDVITSSGLPYGMSVPEKAEAEAGTPGGVERKPLVASTALATESLTLST 492
 256 -----SPLKPYGC-----EECG-----KSYRLIS--274
 493 GTSTAVABGLPTFNKFLMKAVEPKSKADENTPPGSEGSALAGVADSGATPMQSLVY 552
 275 -----LNLHKRHHTGEAKYRCDD-----CGKLPT 299
 553 SLPSMALLTNHLKSGSPFPVYLPGLGASPE--TSKIQOLV-----EKID 597
 300 T--SGNLKRHOVHSGE--KPYOCDYCGRSFSDPTSKMRHLEHTDTEHKPCPHCDKFN 355
 598 ROG-----AAVAS-----TASGAPTTAPAPSSASGPNOCVLCRLVSCP 639
 356 QVGNLKAHLKIHADGPLKRCRGNETTSG--NLKRLRLHSGEKPRVYCVHQRFADP 413
 640 RALRLHYGONGGERPFCVKVCGRAFSTRGNIRAHF--VGHKTSFAARA----- 685
 414 GALAHVPIHTGEKPCQCLICGKAFQASSLIHAVRHDTGKRPVYCERCGRKRFVQSSOLA 473
 686 -----QNSCPICCKFTNAVTLQOHVRLGQIPNGGSLSEGGAAQENSSEQ 725
 474 NHIRHNDIRPKCTVCNKAFAVNVGDLSKHIITF--GEKPLCDKCGRFNRVNDLRSHV 532
 736 STASGSGFPQSQOPSPSEESSEEEDEEEDVDYDEDSLAGRSSESGEKAISVVG 795
 533 KTVH-----QGRKAKILPEPD-----GSEL--NIVTVAS 560
 796 DSEEVSGAEEVATSYAAPPTVKEMDSNEKAPQHTLPPPPPPDNLDPQMEQG--TSDV 854
 561 DDMVTLATEALAAVATVQLTVV-----PVAATAVTADE 592
 855 SGAMEEAKLEGISSMALITOGEGSTPLVEELNLPKMKKDPGESSGRKACEVGOS 914
 593 TEALKAEI-----TAVAKOV-----DEADP-NIOLITACDSCGEK 626
 915 PPTOTALAEHOKTH 928
 627 FLDAISLAQHVRIH 640
 RESULT 24
 Z071_XENLA
 ID Z071_XENLA STANDARD; PRT; 898 AA.
 AC P18751;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE Oocyte zinc finger protein XLCOF7.1 (Fragment).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodidae; Xenopus.
 NCBI_TaxID=83355;
 DB 1 LKSLTLVAESVSESELTELGEKVIIVEGKTARVTOGSDSKROYV----- 48
 238 PLSPPIPAQOTKTTAS--SSSSSSSGAEP--KQAFHLHYPLGSGQPFPSVGVGRSHK 294
 49 -----PNOGKEEPAYATAOPKPEAPDPAKEG-----PAEGDP--GGVDNAE 92
 295 PTPAP-SPALPGSTPQLALASPHLAPFGTGLLAQCLGAARGL-----E 337
 93 ASPAAVSPSRPQPAEVEGN---SSPGEKSDAPST--EARGMELEGKEDEEAWVEDE 147
 338 AAASPLGLKPK-----NGSGELG-----YGEVTSLEKPGGRHKC 372
 148 EAKTPKAAQPKPSEKEMEDNESGSTDSGENSEGSTRLRREGTYSDFRESKAYAAVTHKC 207
 373 RFCAKAVGSDSALQIHLNSHTGERPKVCNVCGRFTTGNLKVPHRHKREKPHVQMNH 432
 208 EDCGKEFTHTGTFKRIHRIHTGEKPKSCRCNKAFSDPAACKAHEKTH----- 255
 433 PVPEHLVDVITSSGLPYGMSVPEKAEAEAGTPGGVERKPLVASTALATESLTLST 492
 256 -----SPLKPYGC-----EECG-----KSYRLIS--274
 493 GTSTAVABGLPTFNKFLMKAVEPKSKADENTPPGSEGSALAGVADSGATPMQSLVY 552
 275 -----LNLHKRHHTGEAKYRCDD-----CGKLPT 299
 553 SLPSMALLTNHLKSGSPFPVYLPGLGASPE--TSKIQOLV-----EKID 597
 300 T--SGNLKRHOVHSGE--KPYOCDYCGRSFSDPTSKMRHLEHTDTEHKPCPHCDKFN 355
 598 ROG-----AAVAS-----TASGAPTTAPAPSSASGPNOCVLCRLVSCP 639
 356 QVGNLKAHLKIHADGPLKRCRGNETTSG--NLKRLRLHSGEKPRVYCVHQRFADP 413
 640 RALRLHYGONGGERPFCVKVCGRAFSTRGNIRAHF--VGHKTSFAARA----- 685
 414 GALAHVPIHTGEKPCQCLICGKAFQASSLIHAVRHDTGKRPVYCERCGRKRFVQSSOLA 473
 686 -----QNSCPICCKFTNAVTLQOHVRLGQIPNGGSLSEGGAAQENSSEQ 725
 474 NHIRHNDIRPKCTVCNKAFAVNVGDLSKHIITF--GEKPLCDKCGRFNRVNDLRSHV 532
 736 STASGSGFPQSQOPSPSEESSEEEDEEEDVDYDEDSLAGRSSESGEKAISVVG 795
 533 KTVH-----QGRKAKILPEPD-----GSEL--NIVTVAS 560
 796 DSEEVSGAEEVATSYAAPPTVKEMDSNEKAPQHTLPPPPPPDNLDPQMEQG--TSDV 854
 561 DDMVTLATEALAAVATVQLTVV-----PVAATAVTADE 592
 855 SGAMEEAKLEGISSMALITOGEGSTPLVEELNLPKMKKDPGESSGRKACEVGOS 914
 593 TEALKAEI-----TAVAKOV-----DEADP-NIOLITACDSCGEK 626
 915 PPTOTALAEHOKTH 928
 627 FLDAISLAQHVRIH 640

```

RN      [1]
RP      MEDLINE=89345612; PubMed=2503827;
RX      Knocheil W., Poeling A., Koester M., el Baradi T., Nietfeld W.,
RA      Boumeester T., Pleler T.;
RT      "Evolutionary conserved modules associated with zinc fingers in
RT      Xenopus laevis.";
RL      Proc. Natl. Acad. Sci. U.S.A. 86:6097-6100(1989).
[2]
RN      SEQUENCE OF 284-898 FROM N.A.
RP      MEDLINE=90040696; PubMed=2509712;
RX      Nietfeld W., El-Baradi T., Mentzel H., Pleler T., Koester M.,
RA      Poeling A., Knocheil W.;
RT      "Second-order repeats in Xenopus laevis finger proteins.";
RL      J. Mol. Biol. 208:639-659(1989).
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
EMBL: M25866; AAA50013.1; .
PIR: A33282; A33282.
HSSP: P08047; 1SP2.
InterPro: IPR000822; Znf_C2H2.
DR      Pfam: PF00096; zf_C2H2; 21.
DR      SMART: SM00355; Znf_C2H2; 21.
DR      PROSITE: PS00028; ZINC_FINGER_C2H2_1; 21.
DR      PROSITE: PS50157; ZINC_FINGER_C2H2_2; 21.
KM      ZINC-finger, Metal-binding; DNA-binding; Repeat.
FT      ZN_FING      289      311      C2H2-TYPE.
FT      ZN_FING      342      364      C2H2-TYPE.
FT      ZN_FING      370      392      C2H2-TYPE.
FT      ZN_FING      398      420      C2H2-TYPE.
FT      ZN_FING      425      448      C2H2-TYPE.
FT      ZN_FING      426      448      C2H2-TYPE.
FT      ZN_FING      454      476      C2H2-TYPE.
FT      ZN_FING      481      504      C2H2-TYPE.
FT      ZN_FING      482      504      C2H2-TYPE.
FT      ZN_FING      510      532      C2H2-TYPE.
FT      ZN_FING      537      561      C2H2-TYPE.
FT      ZN_FING      538      561      C2H2-TYPE.
FT      ZN_FING      567      589      C2H2-TYPE.
FT      ZN_FING      595      617      C2H2-TYPE.
FT      ZN_FING      623      645      C2H2-TYPE.
FT      ZN_FING      651      673      C2H2-TYPE.
FT      ZN_FING      679      701      C2H2-TYPE.
FT      ZN_FING      707      729      C2H2-TYPE.
FT      ZN_FING      735      758      C2H2-TYPE.
FT      ZN_FING      764      786      C2H2-TYPE.
FT      ZN_FING      792      814      C2H2-TYPE.
FT      ZN_FING      820      842      C2H2-TYPE.
FT      ZN_FING      848      870      C2H2-TYPE.
FT      ZN_FING      876      898      C2H2-TYPE.
FT      NON_TER      898
SQ      SEQUENCE      898 AA; 101167 MW; 795D806E5696B0B6 CRC64;
Query Match      5.1%; Score 271.5; DB 1; Length 898;
Best Local Similarity 19.7%; Pred. No. 3.6e-05;
Matches 199; Conservative 95; Mismatches 387; Indels 327; Gaps 41;
OY      27 EHHNPOVAKKCAQFSDPTFLAHONS-CCTDPPVAVIIGQENPNSSSASSAPRPGH 84
OY      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      128 QQLRPOAC-----EYKDESIVTAHMEATLCNSD-----GNFIRPENPEISBGEQ--- 173
OY      85 SRSQVMDTHSNPPDSGSSPPPTWCPERRGESSGQFLVAATGTAAGCGGGLIASPK 144
OY      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      174 -----PPANGIK-----EE-----ATSSSEGN-----QSD 193
```

```

OY      145 LGATPLPESTPAPPPPPPPPGVSGHLNIPL-----LLEELRVLQORIQHQQ 196
OY      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      194 CSINPLTERIGQTDPTP-----IMGCSLNTSLSDNYISDELKEEPTSCGRNOSDS 246
OY      197 MTEQICROYLLIGSQIVGAPASPELPGTAASSTKPLPLPSP----- 242
OY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      247 IT-----PLTEQIQGTDTPTPIMGCSLKDKNKYDCGNPHMSPKNTLRRKYSQNECH 295
OY      243 ---IKPAQTGKTTASSSSSSSGAEPKQAFHL---YHPLSGHPRSVGVGR----- 291
OY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      296 EYLHKRDFGKHQMHKKREKFSCEGKCFNLQCDPRHQTGK-PPSCSKGKCFAPL 354
OY      292 -----SHKPPAPSPALPGSTQDLIASPHLAFPGTTGLLAOCIGARGLIEAA 340
OY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      355 SDLYNRRIRHTGERPSPSCSGKGFTRPMALINHRHTGKPFSCSEC----- 403
OY      341 SPGLLKPNGSELGYEIVISLEKPGRRKRCFCAVFGSDALQILHRSHTGERPYC 400
OY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      404 --GKCFSSKQSS--LVNHQRTHIGEK--FCCSCDKCFASSSELINHQRTHTGKPFSC 456
OY      401 NVCGNRFTTRGNLKVFNHRREKYPHYQMPNHPYRPHLDYITSSGLPYGMVPRPKAE 460
OY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      457 SECQKCFTHMS--HFANHQ-----MIHTG----- 478
OY      461 EAGTPGGGVERKPLVASTALSTESLTLLSTGTAVAPGLPTNKEVLMKAVEPKSKA 520
OY      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      479 -----EKPFCCSKGKCFASSSDL-----TPHR-----RTHH 505
OY      521 DENTPPGSE-GSAIAGVADSGSATRMQL-----SKLVTSLPVSMALLTNHLKSTGS 569
OY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      506 REKTFSCSECGKCF---SNHSHLARHQMHHTGKPFCCSECGKCFSSSGLTAHQORTIH 562
OY      570 PFPVYVLEPLGASPSSETSKL--QQLVEKIDROGAVAVASTAGATTAPARASSSASGN 627
OY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      553 KYKPFSCSAGKCFKSNRSHLIRHOMH-----TGKPP 595
OY      628 QCVICLRVLVSCPRALRLHYGOHGERPFCKVCGRAFSTRGNLRAHFYHKTSPARAQN 687
OY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      596 SCFECRKCFNSPNBLARHQHTHTEGKPFSCSECGKCFASSDLDTH--HRHTGSK-PF 651
OY      688 SCPICQKKTNAVTLQOHVBMILGQIIPNGSALSSEGGAAQENSSEOSTASGPGSFPDP 747
OY      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      652 SCSECGKCYSKSSLVNHQRTH-GEKPFSCSKCDKCFASSSELINHQHTHGEKAF--- 707
OY      748 OSQOQSPREEMSEBEDEBEEDVTDDEDSLARGSESGEKAIVRGSEF--VSG--- 802
OY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      708 -----SCSECGKCFTRNSQLSRHQMHHTGKPFISC-PPCECFVSSQL 750
OY      803 -AEEVAVTVAAPTGYKE---MDSNKAPOHTLPPPPPPDNLDPQMEQTSVSGA 857
OY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      751 TAAQOAHNMVAKPFSCLCEGKCFNSRNSFNARHQMHHTGKPFSCSE---CRKGFSSQSL 807
OY      858 MEEKALEGISPMALTOEGEGTSTPLVEELNLPKMKDGGESSGRKACEVGGSGFPPT 917
OY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      808 ARHQ-----MTNIGE---KPF-----SCSECGKRFESN 831
OY      918 QTALEHQAHTPRKDGPLFTCVPCRGQGLDRAATLKKHMLAHNQVPRFA 965
OY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      832 QSHLARHQMHHTGKPF-FSCSECAKGFSSQSGLANHQ-MTHHTGKPEFA 877
RESULT 25
ZFH1_DROME
ID      ZFH1_DROME      STANDARD;      PRT; 1060 AA.
AC      P28166;
DT      01-OCT-1994 (Rel. 30, Created)
DT      01-OCT-1994 (Rel. 30, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Zinc-finger protein 1 (Zinc-finger homeodomain protein 1).
GN      ZFH-1.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC      Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
```


"The identification and characterization of KRAB-domain-containing zinc finger proteins."

Genomics 12;581-589(1992).

- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.

- SUBCELLULAR LOCATION: Nuclear.

- SIMILARITY: BELONGS TO THE KRUPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).

EMBL: L75847; AAB05653.1; -.
DR EMBL: M67509; AAA36133.1; -.
DR EMBL: M67509; AAA36134.1; -.
DR HSSP: P08046; IAI1.
DR TRANSFAC: T04988; -.
DR GeneW: HGNC:13111; ZNF45.
MIM: 601781; -.
DR InterPro: IPRO01909; KRAB.
DR InterPro: IPRO00822; ZnF_C2H2.
DR Pfam: PF00096; zf-C2H2; 15.
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PR00048; ZINCFINGER.
DR ProDom: PD000003; ZnF_C2H2; 13.
DR SMART: SMO0349; KRAB; 1.
DR SMART: SMO0355; ZnF_C2H2; 15.
DR PROSITE: PS00805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 15.
DR PROSITE: PSS0157; ZINC_FINGER_C2H2_2; 18.
KM Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat; Polymorphism.
FT DOMAIN 8 /78 KRAB.
FT N1 192 662 ZINC FINGERS.
FT N2_N1_FING 192 214 C2H2-TYPE.
FT N3_N2_FING 276 298 C2H2-TYPE.
FT N4_N3_FING 304 326 C2H2-TYPE.
FT N5_N4_FING 332 354 C2H2-TYPE.
FT N6_N5_FING 360 382 C2H2-TYPE.
FT N7_N6_FING 388 410 C2H2-TYPE.
FT N8_N7_FING 416 438 C2H2-TYPE.
FT N9_N8_FING 444 466 C2H2-TYPE.
FT N10_N9_FING 472 494 C2H2-TYPE.
FT N11_N10_FING 500 522 C2H2-TYPE.
FT N12_N11_FING 528 550 C2H2-TYPE.
FT N13_N12_FING 556 578 C2H2-TYPE.
FT N14_N13_FING 584 606 C2H2-TYPE.
FT N15_N14_FING 612 634 C2H2-TYPE.
FT N16_N15_FING 640 662 C2H2-TYPE.
A->T (IN DBSNP:1047452).
/FtId=VAR_012019.
R->K (IN DBSNP:399098) .
/FtId=VAR_012020.
T->A (IN DBSNP:388706) .
/FtId=VAR_012021.
P->R (IN DBSNP:388685) .
/FtId=VAR_012022.
R->K (IN DBSNP:407731) .
/FtId=VAR_012023.
SO SEQUENCE 682 AA; 78242 MW; 9C8E536EBD0FCB8 CRC64;

Query Match 5.18; Score 270; DB 1; Length 682;
Best Local Similarity 22.0%; Pred. No. 3.2e-05;
Matches 132; Conservative 56; Mismatches 189; Indels 222; Gaps 23;

364 EKPGGRHKRCFAKKVFSDSLQIHLSHTGERPKVCNCGNFFTRGNLKVH--FRHR 421
161 ERP---YKGGEHVKSFSWSHLQINQRNHAGCKPYKCCEKDVAEFRRSSLDQHQRVSRA 217

```

QY 442 EKKRHQNMHPHYREHLDYVITS--SGIPQMSVY-----PEAAEEAGAPGGVEKPLV 475
Db 218 KSYTN-----DASYRFSQSRSHLPHQRVPTGENPYVEECGRNVKAS-----260
QY 476 A$T$A$L$A$E$S$L$L$T$G$T$A$V$A$P$G$L$P$T$E$K$F$V$L$M$K$A$V$E$P$K$S$K$A$D$E$N$T$P$P$G$E$G$A$I$A$G 535
Db 261 -----SHQ$Q$A$P$L$I$V$H$T$G-----E$K$Y$K$C$E$-----CG 282
QY 536 VADSGATPMQSLKVLTSLP$M$A$L$L$T$H$L$K--$T$G$S$F$P$P$Y$V$L$E$P$L$G$A$S$E$T$S$K$L$O$O$L$V$E 594
Db 283 V--GFSQR-----SYLQVHLK$V$H$T$G$K--$P$Y$K$E$E$C$G$K$S$F$M$R$S$R$L$Q$----320
QY 595 KIDRQGA$V$A$V$A$S$A$G$A$P$T$T$A$P$A$P$S$S$A$S$Q$P$N$O$C$V$L$C$R$V$L$S$C$P$A$R$L$R$H$Y$O$G$H$E$P 654
Db 321 -----A$H$E$R$H$T$G$E$K$P$Y$K$M$A$C$K$S$T$S$Y$S$H$L$N$I$H$C$H$I$H$T$G$E$K$P 359
QY 655 F$K$C$V$G$A$F$S$T$R$G$N$L$P$A$H$F$V$H$K$T$S$P$A$A$R$A$O$N$S$C$P$I$C$K$P$T$N$A$V$L$O$O$H$V$M$H$L$G$O$I 714
Db 360 Y$C$E$G$C$K$G$F$S$V$S$H$L$Q$A$H$O$I$S$H---$T$G$E$K$P$Y$K$E$E$C$G$K$F$C$R$A$S$N$L$D$H$Q$H$T--G$E$K 414
QY 715 P$N$G$A$L$S$E$G$G$A$A$Q$E$N$S$E$O$S$T$A$S$G$C$S$F$Q$P$O$Q$S$P$E$E$M$S$E$E$E$D$E$E$E$D$Y$T$D 774
Db 415 P$Y$O$C$D$A$C$G$G$F$S$R$S$D$P$N$I$F$H$R$V$H$T$G-----E$K$P$K$E$E-----448
QY 775 E$D$S$L$A$G$S$E$S$G$E$K$A$L$S$V$G$D$E$E$V$S$G$A$E$E$V$A$T$S$V$A$P$T$Y$K$E$M$S$N$E$K$A$Q$H$L$P$P 834
Db 449 ---C$K$G$F$S$Q$A$S$N$L$A$H$Q$H$T$G$E-----469
QY 835 P$P$P$D$M$L$D$P$Q$P$M$E$O$G$T$D$V$S$G$A$E$E$E$K$L$B$G$I$S$S$P$M$A$L$T$O$E$G$E$S$T$P$P$L$V$E$L$N$L$P$E$A 894
Db 470 -----K$P$Y$K$C$G$N$C-----G$K$G$F$S$R$--$S$D$L$N--$H 490
QY 895 M$K$D$P$E$S$S$R$K$A$C$E$V$G$G$S$P$T$O$T$A$L$E$H$Q$K$T$H$K$D$G$L$F$C$V$F$C$R$Q$G$L$D$A$T$L$K$H 953
Db 491 C$R$I$H$T$G$E$K$P$Y$--C$E$R$C$G$K$A$F$S$O$F$S$I$O$V$H$Q$V$H$T$G$E$K$P--Y$O$A$C$E$G$C$G$F$S$V$S$Q$L$A$H 546

RESULT 29
EVIL_MOUSE
ID EVIL_MOUSE STANDARD: PRT: 1042 AA.
AC P14404;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ecotropic virus integration 1 site protein.
GN EVIL OR EVI-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88311086; PubMed=2842066;
RA Morishita K., Parker D.S., Mucenski M.L., Jenkins N.A., Copeland N.G.,
RA Ihle J.N.;
RT "Retroviral activation of a novel gene encoding a zinc finger protein
in IL-3-dependent myeloid leukemia cell lines.",
RL Cell 54:831-840(1988).
CC -1- SUBUNIT: MAY INTERACT WITH CTBP1 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-11 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M21829; AAA40581.1; ALT_INIT.
PIR: A31591; A31591.

```

R HSSP; P08047; 1SP2.
R TRANSFAC; T00273; .
R MGD; MGI:95457; Ev11.
R InterPro: IPR000822; ZnF_C2H2.
R Pfam: PF00096; zf-C2H2; 10.
R PRINTS; PRO0048; ZINCFINGER.
R PRODOM; PD00003; ZnF_C2H2; 4.
R SMART; SM00355; ZnF_C2H2; 10.
R PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
R PROSITE; PS0157; ZINC_FINGER_C2H2_2; 10.
M Zinc-finger; Metal-binding; DNA-binding; Developmental protein;
M Repeat; Proto-oncogene; Chromosomal translocation.
T DOMAIN 21 239 ZINC-FINGERS 1.
T DOMAIN 421 434 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
T DOMAIN 724 803 ZINC-FINGERS 2.
T DOMAIN 877 928 ASP/GLU-RICH (ACIDIC).
T ZN_FING 21 44 C2H2-TYPE.
T ZN_FING 75 97 C2H2-TYPE.
T ZN_FING 103 125 C2H2-TYPE.
T ZN_FING 131 154 C2H2-TYPE.
T ZN_FING 160 182 C2H2-TYPE.
T ZN_FING 188 210 C2H2-TYPE.
T ZN_FING 217 239 C2H2-TYPE.
T ZN_FING 724 746 C2H2-TYPE.
T ZN_FING 752 775 C2H2-TYPE.
T ZN_FING 781 803 C2H2-TYPE.
T SITE 553 557 CTBP-BINDING MOTIF 1 (BY SIMILARITY).
T SITE 584 588 CTBP-BINDING MOTIF 2 (BY SIMILARITY).
T SEQUENCE 1042 AA; 116847 MW; 8DEDF164F536D2FE CRC64;

Query Match 5.1%; Score 269; DB 1; Length 1042;

Best Local Similarity 21.9%; Pred. No. 5.4e-05;

Matches 162; Conservative 72; Mismatches 283; Indels 224; Gaps 31;

367 GGRHCRFCAKVGSGSALQIHLS-HTGERPYKMGVGNRRTRGNLKVHFRHH----- 420
128 GKHTCEKCAKVFPTDSNLRHLSQNHGARAHAACRECGKTRATSGLKHKHHSVKR 187
421 -----REKYPHV-----QMPNH-PVPEHLDVIT 443
188 FICEVCHKSYTQFSNLKRNHMDCRTOIKCKDGMFSTSLKHNRFCEGKNHFAA 247
444 SSGLYGMVPRKAEAA-----GTPG---GGEKPLVASTALSTSLTSLT 492
248 GGFQGGQISLPGPRADKTMVNVSHANPGGLADYFGTNRHP-----A 289
493 GTSYAVAPCLPTFNKFLVLMKAVBEKSKADENTPRSGESALIGVADSGSATRMQLSKLVT 552
290 GLTPRTAAGFSFSGFLFRSGLGNRRPLIRASPP-----VGLSTEGSNKCO-SPULT 342
553 SLPSWALLNLHKSTGSFRRPVYLEPLGASPSSETSKLOLVKIDRQGA---VAVASTAS 609
343 HPQLIRAPQDILKALSKNRPVGNKRVELLERSESE-ERPLEKISQSESSDDLDVSTPS 401
610 GA--PTTAPARASSASGRPGCYICLVLSCPRLRLHNGHGEGRPFKCKVCGRAFSTR 667
402 GSDLETTSGSDLESDESE-----CKEKNGMKR-- 432
668 GNLRHNFHGHKTSAPARAONSCPIOCK-----FTNAVTLQOHVHMHGGLDIPNGSGA 720
433 -----DKVSP---LQVLASTNKKENHNHSHVFSASVEBQSAV---SGAVNDSIKA 476
721 LSE-----GGGAOENSSSEOSTASGPGSPPOP-----OSQOQSPSE----- 755
477 IASIAEKYFGSGVLQDKKVGALPYRSMFPLPFPPAFSGSMYPPRPDLRLSLPLKMP 536
756 ---EEMSEEEDEEEDVT---DEDSLARGSGESGGEKAIS-----VRGDSSEVS 801
537 QSPSEYKRLQKGSSESPFLDITRKDKDEPLTSGPSKPGTPTASQOQPLDLSMGSRGAS 596
802 GAE--EEVATSTAAPTIVKEMD---SNEKAPQHILPP----- 834
597 GTKLLEPRKNHVGEEKKSGMMDTRPSSDGLQHAHPTPEFMDPIYRVERKRLDPLRLALK 656

QY 835 ---PPPDNLDPMPMEQGTSPDVSGAMEEAA-KLEGISS--PMAA-LTQEGEGTSPPLV 886
Db 657 EKYLRPSGFLFRHQ-----MSAIEENAEKLESALKPEASELQSVSMFSFRA 707
QY 887 EELNLEPAKKDPGESSGRRACEVGSGFPOTALEEHQKTHPKDPLFTVCYRCGFELD 946
Db 708 PPNLTLPENLIRKGE---RYTCRCYCKIFPRSANLTRLRHTGTGEQ-YRCKYCDRSFSI 763
QY 947 RATLKKHMLAHQVPPRPAH 967
Db 764 SSNLDQHVANNHKKERFKCH 784

RESULT 30

PRDD_HUMAN STANDARD; PRT; 717 AA.
ID PRDD_HUMAN
AC Q9H4Q3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE PR-domain zinc finger protein 13.
GN PRDM13 OR PFM10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yang X.-H., Huang S.;
RT "A family of novel PR-domain (PRDM) genes as candidate tumor suppressors".
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear (potential).
CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AY004253; AAG13448.1; .
DR Genew; HGNC.13998; PRDM13.
DR InterPro; IPR001214; SET.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF00096; zf-C2H2; 4.
DR SMART; SM00317; SET; 1.
DR SMART; SM00355; ZnF_C2H2; 4.
DR PROSITE; PS0280; SET; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 1 126 SET.
FT ZN_FING 147 169 C2H2-TYPE.
FT ZN_FING 583 605 C2H2-TYPE.
FT ZN_FING 611 633 C2H2-TYPE.
FT ZN_FING 640 663 C2H2-TYPE.
FT DOMAIN 311 317 POLY-ALA.
FT DOMAIN 362 373 POLY-HIS.
FT DOMAIN 382 388 POLY-PRO.
FT DOMAIN 564 575 POLY-GLY.
SQ SEQUENCE 717 AA; 75059 MW; FB45135P6488254A CRC64;
Query Match 5.1%; Score 268.5; DB 1; Length 717;
Best Local Similarity 24.5%; Pred. No. 3.9e-05;
Matches 173; Conservative 55; Mismatches 262; Indels 215; Gaps 34;
QY 132 AGGGGGILT-----ASPKLGATPLPESTPAPP-----PPPPPPGVGSG 173

Db 173 SGGGGAGFLHNEHARAGAVPAADGLGLSPKPPAPDPAPASQACTLRNHLGPP----- 227
OY 174 HLNPLILEELRVLOQRQHQMTFOICROVILLGLGOTVGAPASSELPGTAASST 233
Db 228 -----VQAC-----GAREGIKREASSAP-----SAT 248
OY 234 KPLPLPSPKPAOTGKTATSSSSSSSSSGAEPKQAFPHLYHPLGSHQHPFVGGVGRSH 293
Db 249 SP-----TPGKWGQKKKKEQIDRALDMSGARGQHFLGI-----VGG----- 287
OY 294 KPTPAPSPALPGSTQDLIASPHLAF-PGTTGLLAQCLGARGLEAASPGILKPKNGSG 352
Db 288 -----SSAGVS-----LAFYPGVRS-AFKPAGLARA-AAAHQDPYHESSS 328
OY 353 ELGYGEVYISLEKPECGRIKRCRCAKVFGSDSALQIHLRSHTGEPYKCNVCNRTTTCGN 412
Db 329 KQAGLALGRL--LGGGRACG-----RPGSGEN-----SAAAG 359
OY 413 LKVPHRHREKYPHYOM-----NPHVPEHLDYVITSSGLPYGMSVPEKAEEAGTPGG 467
Db 360 AGHHHHHHHHHHHHKCLLAGDRPPPPPEGLPCSGALRGFPL-LSVPEEA-----SAFK 413
OY 468 GVERKPLVASTTALSTESLTLSTGTSTAVABGLPTFNKFLMKAVEPKSKADENTPPG 527
Db 414 HVERAPRAAA--ALPGARVAQL-----PPAPGLP-----LEGCALPLPDGGLKATYRG 459
OY 528 SEGSAIAGVADSGSATRMQLSKLVTSPLSMALLTNHLKSTGSPFP-----VYLEPLGA-- 581
Db 460 GECCHLPVMPAFVYNGEL--LYGSPATAYVPLKLHFGGLKYPEGISYFSGPAAAL 517
OY 582 SPSESKL-----QOLVE-----KIDROGAVAVASTASGAPTSAPAPSS 622
Db 518 SPALGSLASIDRELTAMHQOLSEMAAGKGRGLD-SGLTPRAVAAAGG-TGGGSGSG 575
OY 623 ASGP--NQCVCILRVLSCPRALRLHYGHGERPCKYCGRAFTSTGNLRAHFVGHKT 679
Db 576 AGKPRGTGILCYGKLYSKYGLKIMHRTHTGKPKCKVCLRPGRDPSNLKHLRLH-- 633
OY 680 SPAAANQSCPTCQKKTAAVTLQOVNRMHLGGQIPNGSALSEGGAQAQENSSEOSTAS 739
Db 634 -AEGNTPYRCFCGKVLVRRLERHNVKSRHPCG-----SLAKAG-----D 674
OY 740 GPGSPPOSOQOPSPREESEEEDEEDVDEDSLAGRSE 784
Db 675 GPGAP-----GYPPECDPKSDSDVDVCFITDQSDPE--VGGGE 714
RESULT 31
Y441_HUMAN STANDARD: PRT: 697 AA.
ID Y441_HUMAN
AC 043167;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein KIAA0441.
CN KIAA0441.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=Brain;
RC MEDLINE=98116655; Pubmed=9455477;
RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura O.,
RA "Prediction of the coding sequences of unidentified human genes. VIII.
RT 78 new cDNA clones from brain which code for large proteins in
RT vitro.";
RL DNA Res. 4:307-313(1997).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).

CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL: AB007901; BAA23713.1; -
DR HSSP: P08046; 1AIG.
DR InterPro: IPR000637; AT_hook.
DR InterPro: IPR000210; BTB_POZ.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF00096; zf-C2H2; 8.
DR Pfam: PF00651; BTB; 1.
DR Pfam: PF02178; AT_hook; 1.
DR PRINTS: PR00048; ZINC_FINGER.
DR PRODOM: PD000003; ZnF_C2H2; 2.
DR SMART: SM00384; AT_hook; 1.
DR SMART: SM00225; BTB; 1.
DR SMART: SM00355; ZnF_C2H2; 8.
DR PROSITE: PS00097; BTB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 8.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT DOMAIN 10 133
FT ZN_FING 294 512 ZINC_FINGERS.
FT ZN_FING 294 316 C2H2-TYPE.
FT ZN_FING 322 344 C2H2-TYPE.
FT ZN_FING 350 372 C2H2-TYPE.
FT ZN_FING 378 400 C2H2-TYPE.
FT ZN_FING 406 428 C2H2-TYPE.
FT ZN_FING 434 456 C2H2-TYPE.
FT ZN_FING 462 484 C2H2-TYPE.
FT ZN_FING 490 512 C2H2-TYPE.
SQ SEQUENCE 697 AA; 78292 MW; F2BD3C144626544 CRC64;
Query Match 5.1%; Score 268; DB 1; Length 697;
Best Local Similarity 21.0%; Pred. No. 4e-05;
Matches 115; Conservative 68; Mismatches 169; Indels 196; Gaps 20;
OY 364 EKPQG-RHKRCFCAGVFGSDSALQIHLRSHTGEPYKCNVCNRTTTCGNLNVHHRRE 422
Db 287 KRPQGEARCKDCGKVEKYVNHFLAIHQRSHTGERPFCNCGKGPAAKSHLSQVHTMHTG 346
OY 423 KYPHYQMNHPVPEHLDYVITSSGLPYGMSVPEKAEEAGTPGGVEKRPVASTALS 482
Db 347 ERPYT-----CTVCSKAL-----TTKHS 364
OY 483 ARESLTLSTGTSTAVABGLPTF-----NKFVLMKAVEPKSKADENTPPGSESAIAGVA 537
Db 365 LLEHMSLHS-----GKSTFCDCGKTFQSQR-QLKSYRYHT-----CHSLPECK 409
OY 538 DSGSATRMQLSKLVTSPLSMALLTNHLKS-TGSEPPVYLEPLGASPSSETSKLQLEVKI 596
Db 410 DC-HRKFMVDSQ-----LKKHLRTHGKEKPF----- 434
OY 597 DRQGAVAVASTASGAPTSAPAPASSASGPNQCVLCVLSCPRALRLHYGHGERPCK 656
Db 435 -----TCEICGKSFYAKSSLSQTHIRIHGRKPY 463
OY 657 KVCYGRAFTSGNLRAHFVGHKTSPAAANQSCPTCQKKTAAVTLQOVNRMHLGQIIPN 716
Db 464 GCITGKSFSDSAAKRRICILH-----TGKKPSCPCNQLQFALDNLKALKIKHSREKHA 519
OY 717 GGSALSEGGAQGE-----NSSEOS-----TA 738
Db 520 DASSIS-GSSNTEVEVRNIILOLPYQLSTSGEQEIQLVTDVSHNINFMPSGISIVTA 578

```

Y      739  SGPSFPPQ-----SOQPSPPEEM-SPEEEDDEEDVDVDEDLAGSGSGGKKA   790
          :|         |||         ::||         |::|         |::|
b      579  ESSQMNTADQANLTLTITQQPDELQNLIIISAQEQETHIOSLNMTESOMG-PSOTEPVNV   637
          :|         |||         |||         |||         |||
Y      791  ISVGDSSEVSQAEE-----VATSVAAPTYTKENDSNKRAQHTLP PPPPPNDLDPH   844
          :|         :|         :|         :|         :|
b      638  ITLSKETLEHLHAHQEQTEBELHILATSTSDPAQLDLTLQE-----GCPRRPHNHVPQ   689
          :|         :|         :|         :|         :|
Y      845  QPMQDGT 852
          :|         :|
b      690  TPLGOEQS 697

RESULT 32
192_HUMAN
D_2192_HUMAN STANDARD: PRT; 578 AA.
C_Q15776; Q9H4T1;
T_16-OCT-2001 (Rel. 40, Created)
T_16-OCT-2001 (Rel. 40, Last sequence update)
T_15-JUN-2002 (Rel. 41, Last annotation update)
T_Zinc finger protein 192 (LD5-1).
N_ZNF192.
S_Homo sapiens (Human).
S_Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C_Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
C_NCBI_TaxID=9606;
X[1]
P_SEQUENCE FROM N.A., AND VARIANT LEU-163.
P_TISSUE-Ovary;
C_MEDLINE=97386587; Pubmed=9244436;
X_A Lee P.L., Gelbart T., West C., Adams M., Blackstone R., Beutler E.;
X_P "Three genes encoding zinc finger proteins on human chromosome 6p21.3:
X_T members of a new subclass of the Kruppel gene family containing the
X_T conserved SCAN box domain.";
X_L Genomics 43:191-201(1997).
N[2]
P_SEQUENCE FROM N.A.
P_Williams S.;
A_Submitted (SEP-2001) to the EMBL/genbank/DBJ databases.
L[3]
P_PARTIAL SEQUENCE FROM N.A.
C_TISSUE-Ovary;
C_MEDLINE=96230927; Pubmed=8673473;
X_X Beutler E., Gelbart T., West C., Kuhl W., Lee P.;
X_P "A strategy for cloning the hereditary hemochromatosis gene.";
X_L Blood Cells Mol. Dis. 21:206-216(1995).
T_-- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
T_-- SUBCELLULAR LOCATION: Nuclear (Potential).
T_-- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
T_FINGER PROTEINS.
C_-- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
C_-- SIMILARITY: CONTAINS 1 SCAN BOX.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outpost at
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: U57796; AAB02260.1; -.
EMBL: AL358933; CACSI5901.1; -.
EMBL: UB8080; AAC51656.1; -.
EMBL: UB8079; AAC51656.1; JOINED.
RSSP: P07248; IPAA.
GeneW: HGNC:12983; ZNF192.
MIM: 602240; -.
InterPro: IPR001909; KRAB.
InterPro: IPR003309; Treg_SCAN.
InterPro: IPR000822; ZnI_C2H2.
Pfam: PF00096; zf-C2H2; 9.

```

DR	Pfam: PF01352; KRAb: 1.		
DR	Pfam: PF02023; SCAN: 1.		
DR	PRINTS: PRO0048; ZINC_FINGER.		
DR	ProDom: PD000003; ZnF_C2H2; 8.		
DR	SMART; SM00349; KRAb: 1.		
DR	SMART; SM00431; LER: 1.		
DR	SMART; SM00355; ZnF_C2H2; 9.		
DR	PROSITE; PS50805; KRAb: 1.		
DR	PROSITE; PS50804; SCAN_BOX: 1.		
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.		
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 9.		
KW	Transcription regulation; zinc-finger; Metal-binding; DNA-binding; Nuclear protein; Repeat; Polymorphism.		
FT	DOMAIN	51	133
FT	DOMAIN	220	316
FT	DOMAIN	322	568
FT	ZN_FING	322	344
FT	ZN_FING	350	372
FT	ZN_FING	378	400
FT	ZN_FING	406	428
FT	ZN_FING	434	456
FT	ZN_FING	462	484
FT	ZN_FING	490	512
FT	ZN_FING	518	540
FT	ZN_FING	546	568
FT	VARIANT	163	163
FT	CONFLICT	34	34
FT	CONFLICT	113	113
SO	SEQUENCE	578 AA; 65757 MW;	5014F5439F4B8AB8 CRC64;

```

Query Match Summary      5.0% ; Score 267; DB 1; Length 578;
Best Local Similarity   22.0% ; Pred. No. 3,7e-05;
Matches    131; Conservative    56; Mismatches    195; Indels    214; Gaps    19;

QY 177 IDLIEELRYLOOROHOMONTEQC-----ROYLLIGS-----ICGTVGAP 218
Dy 100 ILTLEPELTLYVD--HOLENGEEVVLLEDLERIDILGRPVSAHVHGRIWEEVYHS 157
QY 219 ASPSELPGT--GAASSTKPLLPFSPIPKAQGTGTTAASSSSSS--SSG-- 263
Db 158 ASAREPNTQLOSEATQHKSVPPOESQRAMSTSSPTPSOKSGSSGDQENTATILLTGFO 217
QY 264 -----ABPKOAFHLHPRLGSQHPFVSQGVGGSHKTPAPSPALPG 305
Db 218 TLEKIEDMAVSLIREEWLDPSCDKLCDNPNENRNMFSLGGETSESENRELA----- 270
QY 306 STDODLIAS---PHLAPFGTGLLAAOCIG-AARGFEAASPGLTKRKNSGCLGYGEVIS 361
Db 271 -SKOYISTGIQP-----GETAKCKGDVIYRGHEHEARDL-----G 307
QY 362 SLEKPGG-----RHKCRFCAKVYGSDPALIHLHSHTGERPYKCNCGNRRFTRGMLKV 415
Db 308 RLEORGNGPTQERRHKDECGKSFQSSGLVRHWIRHTGEPKYQCNVCQKAFYSALLS 367
QY 416 FHHRREREKPRHYQMNHPRVEHLDYVITSSGLPYGMSPVERKAEEAGTGGGVKERPLY 475
Db 368 HODIHKK-----VKRNICK 381
QY 476 ASTTALSATESLTLSTGSTAVAPGLPTFNKFVLMKAVERPSKADENPPGSEGSAIA 535
Db 382 ECGKAFSQWTGL-----LHQRHTHTEKPRYOQNCQCGKAFQSAG 420
QY 536 VADSGSATMQLSKLVTSLSFWALLTNHLKSTGSPFPFVLEPLGASPSSTSKLOOLVER 595
Db 421 L-----ILHORINISGER--PYECNECGKAFSHSHL----- 449
QY 596 IDRQGAVAASTAGAPTSPAPAPSSASGCNOVCYLRLVLCRPLRLIYYGOHGGERPF 655
Db 450 IGHQNL-----HNGEKRYEDECCKTFPRRSSHLIGHQRHTGKRPY 490
QY 656 KCKVGRAVESTRGNLRAHVEVGHKTSPARAANOSCPICOOKFTNAVTLOOHVHRMLG 711

```

Db 491 KCNCGRAFSOKSGL-----IEHQRIHTGERPYKCKECCGAFNGTGLIQLHRIHTG 542

RESULT 33

ID 2263_HUMAN STANDARD: PRT: 683 AA.

AC 014978; 043387; (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Zinc finger protein 263 (Zinc finger protein FPM315).

GN ZNF263 OR FPM315.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI_TaxID=9606;

SEQUENCE FROM N.A.

MEDLINE=97398134; PubMed=9256059;

Yokoyama M., Nakamura M., Okudo K., Matsubara K., Nishi Y., Matsumoto T., Fukushima A.;

Isolation of a cDNA encoding a widely expressed novel zinc finger protein with the Lef and KRAB-A domains.

Biochim. Biophys. Acta 1353:13-17(1997).

(2)

SEQUENCE FROM N.A.

Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E., Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S., Goodwin L., Bryant J., Tesmer J., Melnick L., Longmire J., White S., Ueng S., Tatum O., Campbell C., Fawcett J., Maltbie M., Mista M., Deaven L.;

Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.

-1- FUNCTION: MIGHT PLAY AN IMPORTANT ROLE IN BASIC CELLULAR PROCESSES AS A TRANSCRIPTIONAL REPRESSOR.

-1- SUBCELLULAR LOCATION: Nuclear (By similarity).

-1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS, SPLEEN, THYMUS, PROSTATE, TESTIS, OVARY, SMALL INTESTINE, COLON AND LEUKOCYTE.

-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

-1- SIMILARITY: CONTAINS 1 SCAN BOX.

-1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: D88827; BAA21853.1; .

EMBL: AC004232; AAC24490.1; .

HSSP: P08045; 12NF.

Genew: HGNC:13056; ZNF263.

MIM: 604191; .

InterPro: IPR001909; KRAB.

InterPro: IPR003309; Treg-SCAN.

InterPro: IPR008222; ZnfC2H2.

Pfam: PF00096; zf-C2H2; 9.

Pfam: PF01352; KRAB; 1.

Pfam: PF02023; SCAN; 1.

PRINTS: PR00048; ZINC_FINGER.

PRODom: PD000003; Znf_C2H2; 9.

SMART: SM00349; KRAB; 1.

SMART: SM00431; LER; 1.

SMART: SM00355; ZNF_C2H2; 9.

PROSITE: PS50805; KRAB; 1.

PROSITE: PS50804; SCAN_BOX; 1.

PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.

PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.

Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein; DNA-binding; Repeat; Repressor.

FT DOMAIN 41 123 SCAN BOX.

FT DOMAIN 217 289 KRAB.

FT ZN_FING 378 681 ZINC_FINGERS.

FT ZN_FING 378 400 C2H2-TYPE.

FT ZN_FING 434 456 C2H2-TYPE.

FT ZN_FING 462 484 C2H2-TYPE.

FT ZN_FING 490 512 C2H2-TYPE.

FT ZN_FING 518 540 C2H2-TYPE.

FT ZN_FING 575 597 C2H2-TYPE.

FT ZN_FING 603 625 C2H2-TYPE.

FT ZN_FING 631 653 C2H2-TYPE.

FT ZN_FING 659 681 C2H2-TYPE.

FT CONFLICT 118

SEQUENCE 683 AA; 77240 MW; 2DBFE6C1FC6793A CRC64;

Query Match 5.0%; Score 267; DB 1; Length 683;

Best Local Similarity 21.0%; Pred. No. 4.3e-05;

Matches 145; Conservative 67; Mismatches 226; Indels 254; Gaps 29;

QY 142 SPKLGATPLPESTAP-----PPPPPP-----PPPG-VSGHLMNI 177

DB 153 SPSEKLEPETERSPQRLQELLGSPQDPQAVKERALSAPWLSLFPPEGMEDEKMTG 212

QY 178 PLILEELR-----VLQQRQIQHQMTEQICROYLL-----LGSIGQTVGAPASPS-ELPG 226

DB 213 POLPESLEDVAMKISQEBMGHODPSKRALSRDVOESENVDLSHIT-----PSQEVFG 267

QY 227 TGAASSTPLPLFSPDKPAQGTGTTAASSSSSSSGAEPKQAFHL-----YH 276

DB 268 TVQGGGK-----LMDP-----SVQCKEGLSPRAPGEKFEENLEGVPSVSENH 315

QY 277 P-----LGSHPFVGVGVGRSHKPT-----APSPALPGSTDLIASHLAFTGTGLA 326

DB 316 PVLPLDARGGVPMSP-PELGRPHDRSQGDWAPP--EGGEOALA----- 358

QY 327 AQCIGAANGLEMAASGGLPKP-----NSGELGYGEV-----ISSE 364

DB 359 ----GASSGRE-LGRKELQPKKHLCLPGCKNSNNILRHQRINHAERLCMGVDCE 413

QY 365 KPGGR-----HKCRFCAKVGSDSALOILHRSHTGERPYKCNVCGNFTT 409

DB 414 IFGGNPRELSLRAHLGEBANHCLECGKCFSONTHLRHQRHTHGEKPYOCNICGKCFSC 473

QY 410 RGNLKVHHRHREKYPHYQMNPRVPEHLDYVITSSGLPVGMSVPRPEKAEERAGTGGV 469

DB 474 NSNL-----HHQRHTHEKPYKCECGEITFAHSSNLRHQR-----HT 513

QY 470 ERKPLVASTTALSTESTLTLSTGTSTAVAPGLPTFNKFLVLMKAVERPKSKADENTPGE 529

DB 514 GERPYKCECGKSFSSSHLV-----IHERTHERELYPSE 550

QY 530 -GSAIAGVADSGSATRMQLSKVLTSIPSMALLTNHKLKSTGSPFPYVLEPLGASGEFTSK 588

DB 551 CGEAVS-----DSTPELTNNH-----GAHRAE--- 571

QY 589 LQQLVEKIDRQAVAVASTASGAPTTSAPAPSSASGPNOCVLCRLVSLCPRALHVGQ 648

DB 572 -KKLFE-----CLTGKSPRQGMHLRHQR----- 596

QY 649 HGERPFKCKVCGRAFTSGNLRANFVGHKTSPPARAQNSCPICQKKTNTVTLQOHVAM 708

DB 597 HTGERPYKCTLCGENBSHNSNL-----IRHQRHTHGEKPYCHCEGDSFSHNSNRIRHLRT 652

QY 709 HLGQIPIPGGSAISFEGCGAAGKSSQSTASG 740

DB 653 HT-GERPYKCECGESFSSSRSLMSHQRHTTG 683

RESULT 34

KRUH_DROME STANDARD: PRT: 845 AA.

ID KRUH_DROME

AC P08155; Q9VMH5;

DT 01-AUG-1988 (Rel. 08, Created)

16-OCT-2001 (Rel. 40, last sequence update)
15-JUN-2002 (Rel. 41, last annotation update)
Kruppel homologous protein 1.
KR-H1 OR KR-H OR CG9167.
Drosophila melanogaster (Fruit fly).
Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A., DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
STRAIN=Oregon-R; TISSUE=Embryo, larva, and pupae;
MEDLINE=20237577; PubMed=10772791;
"Kruppel homolog, a stage-specific modulator of the prepupal ecdysone response, is essential for Drosophila metamorphosis."
Dev. Biol. 221:53-67(2000).
[2]
SEQUENCE FROM N.A. (ALPHA AND BETA ISOFORMS).
STRAIN=Berkley;
MEDLINE=20196006; PubMed=10731132;
Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazee R.C., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Ahl J.F., Agbayani A., An H.-U., Andrews-Pfankuch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Diez S.M.,
Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
Houston D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
Jatani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of Drosophila melanogaster."
Science 287:2185-2195(2000).
[3]
SEQUENCE OF 289-367 FROM N.A.
MEDLINE=87051757; PubMed=3096579;
Schub R., Atcher W., Gaul U., Cote S., Preiss A., Maier D.,
Seifert E., Nauber U., Schroeder C., Kemler R., Jaekle H.;
"A conserved family of nuclear proteins containing structural elements of the finger protein encoded by Kruppel, a Drosophila segmentation gene."
Cell 47:1025-1032(1986).
[4]
-1- FUNCTION: PLAYS A GENERAL ROLE IN THE HIERARCHIES OF GENE
EXPRESSION LEADING TO METAMORPHOSIS.
-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, ALPHA AND BETA (SHOWN HERE); ARE
PRODUCED BY ALTERNATIVE SPLICING.
-1- DEVELOPMENTAL STAGE: BETA ISOFORM IS EXPRESSED DURING

CC EMBRYOGENESIS, MOST ABUNDANT IN MIDEMBRYOGENESIS, AND IN ADULTS.
CC ALPHA ISOFORM IS EXPRESSED FROM EMBRYOGENESIS TO 8 HOURS AFTER
CC PUPARIATION. MAJOR PERIOD OF EXPRESSION IS DURING SECOND INSTAR.
CC SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC -----
CC EMBL: AJ005440; CAA06543.2; -
CC EMBL: AJ005441; CAA06544.2; -
CC EMBL: AE003612; AAF52343.2; -
CC EMBL: AE003612; AAG22417.1; -
CC EMBL: M14940; AAA28660.1; -
CC HSSP: P08047; 1SP2.
CC PDBase: PDB0028420; Kr-h1.
CC InterPro: IPR000822; Znf_C2H2.
CC Pfam: PF00096; Zf-C2H2; 8.
CC ProDom: PD000003; Znf_C2H2; 4.
CC SMART: SM00355; Znf_C2H2; 8.
CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
CC PROSITE: PS0157; ZINC_FINGER_C2H2_2; 8.
CC ZINC-finger, Metal-binding; DNA-binding; Repeat; Alternative splicing;
CC developmental protein.
CC FT DOMAIN 108 148 GIN-RICH.
FT DOMAIN 482 515 SER/THR-RICH.
FT DOMAIN 538 565 SER/THR-RICH.
FT DOMAIN 194 463 ZINC-FINGERS.
FT ZN_FING 194 216 C2H2-TYPE.
FT ZN_FING 271 293 C2H2-TYPE.
FT ZN_FING 299 321 C2H2-TYPE.
FT ZN_FING 327 349 C2H2-TYPE.
FT ZN_FING 355 377 C2H2-TYPE.
FT ZN_FING 383 407 C2H2-TYPE.
FT ZN_FING 413 435 C2H2-TYPE.
FT ZN_FING 441 463 C2H2-TYPE.
FT VARSPLIC 1 54 MISSING (IN ISOFORM ALPHA).
FT CONFLICT 313 313 N -> D (IN REF. 3).
SQ SEQUENCE 845 AA; 91451 MW; A4D878E98DCB372 CRC64;

Query Match 5.0%; Score 266; DB 1; Length 845;
Best Local Similarity 18.2%; Pred. No. 5.9e-05;
Matches 173; Conservative 87; Mismatches 328; Indels 362; Gaps 31;

QY 163 PPP-----PPGVSGHLNT-----PLIEELRVIQ 188
DB 80 PPLPLATTTVSGVGLVPSGQGEHFFELLPQQRQMLQLODHOQEQGFVYQALIQ 139
QY 189 QROTHOMQTEOICROVLLISLGQIV-----GAPASPELPGGAASSTRKLLPLFSP 242
DB 140 HQKQOQOQOQHESITNAPTAAPSQAQRKITEPVGGEFPAASAAVVSQYKRPASKP-----Q 193
QY 243 IKPAQGTKTASSSSSSSSSGAEPKQAF-FLYHPLISQHPFSGVGRSHKPTAPSP 301
DB 194 FKQDQGMFTGSKSARTSHTKSHSKNODLSNGASGAVAPVSTRAILINDAGLPVGP 253
QY 302 ALPGSTDLIASPHLAFPTTGLLAQCLGAARGLAASAPGLKPKNSGELGYEYIS 361
DB 254 KSP--TKRLAV-----AAGADPYC-INVQKTFAPVAPRLIRHYRTHGERPF----- 299
QY 362 SLEKPGRHKRCFAVFGSDSLQILHRSITGTERPYKCNVCGNFTTGNLKVHHRHR 421
DB 300 -----ECCERKHLFVKENLQVHRIRHTKERPYKCDVCGRAFEHSGKLRHMRHT 350
QY 422 EKYPVQNMHPVPEHLIDVYITSSGLPYGMSVPEKAEAEACTPGGVERKPLVASTAL 481
DB 351 GERPH----- 355

QY	482	SAESTSTLLSTGCTAVAPGLPTFNKFKLKAVERPSKKADEENPPGSESGALINGVADSGS	541
Db	356	-----KSVCEKT-----FIOSGQ	369
QY	542	ATRMQLSKLVTSLPSMALITNLHLS--TGSPFPYVLEPLIGASPSSTKLOLVEKIDROG	600
Db	370	-----LVIHRTHTGKKPY-----	383
QY	601	AVAVASTAGAPITTSAPAPSSSASGPNQCYICLRVLSCPALNLTGQHGGEPFKCYC	660
Db	384	-----KCPPEG-----GKGFTGSKQLVHSRTHTGKKPYHCDIC	418
QY	661	GRAFSTRGNLRAHFVGHKTSPPAARAONSCPTCKKFTNAVTLQOHVNHNGGQIPN----	716
QY	419	FRDFGVNHLVTLKLRVQHYGSKCYK-----CTICDETFRKKMEAMHKGH-ANEVPRDEAE	473
QY	717	-----GGSALSEC--GGAQOENSSSEDSTVSGPSPPOP-QSQPP-----	752
Db	474	AAAAAASASTSAGSSAGSPBLQGVSSNSSESSNHSPPSPATKKPRQAPRASKTVAAAT	533
QY	753	-----SPE-DEMSSEEDDEEDEDVTDDESLAGRSGSESGEALISV	793
Db	534	LSTPSPSPSSPSSLSSTSPSSASMSAPPTSAHYLPVQMEADALSDSCVSSAQAAPHST	593
QY	794	RGDSEEVSGAEEEVATVSAAPTVMKEMDSNEKARQHTLPRLPPRPNNLD--HPQPMEOGTS	852
Db	594	YADEE-----PTDLSMQOVQGLPESTVDDYQAAPSLLELQPP-----	632
QY	853	DVSGAMEEAKLEGISPMNALQOBEEGSTPLVEELNLPEAKMKDQGESGKKACEYCG	912
Db	633	--AGLTINALLLEA-----ASIKRRHDDNDQOVEDVHAAQOM-----MOLRGHG	678
QY	913	QSEPTQTALAEHQ---KTHPKDGLPTFCVFCRQGLDRAATLKKHMLAHNQVPPFAPHG	968
Db	679	SLEPTQAPAPSHQPPVPTLHVSD-----LAANYDD--THEATVLIENFKRGDLARHG	728
QY	969	-----PONIAITLSLVPGCSSSIIPSPGLSPFRKDDPTM	1001
Db	729	LHKGVAPVKRESALPNDPVVRVVEAALIGLRSTESPERSSSESDSLMW	778
RESULT 35			
RC	2N43_HUMAN	STANDARD:	PRT: 803 AA.
RC	P17038; P28160; Q96DC1;		
RC	01-AUG-1990 (Rel. 15, Created)		
RC	01-DEC-1992 (Rel. 24, Last sequence update)		
RC	15-JUN-2002 (Rel. 41, Last annotation update)		
RC	zinc finger protein 43 (zinc protein HTF6) (zinc finger protein K0X27).		
RC	ZNF43 OR ZNF39 OR K0X27.		
RC	Homo sapiens (Human).		
RC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
RC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
RC	NCBI_Taxid=9606;		
RC	[1]		
RC	SEQUENCE FROM N.A.		
RC	TISSUE=Cell;		
RC	MEDLINE=91279444; PubMed=1711675;		
RC	Lowering R.; Trowsdale J.;		
RC	"A gene encoding 22 highly related zinc fingers is expressed in		
RC	Lymphoid cell lines";		
RC	Nucleic Acids Res. 19:2921-2927(1991).		
RC	[2]		
RC	SEQUENCE FROM N.A.		
RC	TISSUE=Lymph;		
RC	Strusberg R.;		
RC	Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.		
RC	[3]		
RC	SEQUENCE OF 38-190 FROM N.A.		
RC	MEDLINE=91219421; PubMed=2023909;		
RC	Bellefroid E.J.; Poncellet D.A.; Lecocq P.J.; Revelant O.;		
RC	Marital J.A.;		

RT	"The evolutionarily conserved Kruppel-associated box domain defines a
Rt	subfamily of eukaryotic multifingered proteins.";
RL	Proc. Natl. Acad. Sci. U.S.A. 88;3608-3612(1991).
RN	[4]
RP	SEQUENCE OF 476-531 FROM N.A.
RC	TISSUE=Lymphoid;
RX	MEDLINE=91145339; PubMed=2288909;
RA	Thiesen H.-J.;
RT	'Multiple genes encoding zinc finger domains are expressed in human T
cell';	
New Biol.	. 2:363-374(1990);
- FUNCTION:	MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
- SUBCELLULAR LOCATION:	Nuclear (Probable).
- TISSUE SPECIFICITY:	T AND B CELL LINES.
- SIMILARITY:	BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.	
- SIMILARITY:	CONTAINS 1 KRAAB DOMAIN.
This SWISS-PROT entry is copyright. It is produced through a collaboration -	
between the Swiss Institute of Bioinformatics and the EMBL outstation -	
the European Bioinformatics institute. There are no restrictions on its	
use by non-profit institutions as long as its content is in no way	
modified and this statement is not removed. Usage by and for commercial	
entities requires a license agreement (See http://www.isd-sib.ch/announce/ or send an email to license@sib-sib.ch). -----	
EMBL:	X59244; CAA41932.1; -
DR	EMBL:
DR	EMBL:
DR	EMBL:
DR	PIR:
DR	PIR:
HSSD:	P08048; 7ZNF-
TRANSFAC:	T04986;-
GeneW:	HGNC:13109; ZNF43.
MIM:	603972;-
InterPro:	IIPRO01909; KRAAB.
Dr Interpro:	IIPRO00822; Znfc_C2H2.
D Pfam:	PF00096; zf-C2H2; 21.
D Pfam:	PF01352; KRAAB; 1.
DR PRINTS:	PR00048; ZINCFINGER.
DR PRODOR:	PD000003; Znfc_C2H2; 16.
DR SMART:	SMD0349; KRAAB; 1.
DR SMART:	SMD0355; Znfc_C2H2; 21.
DR POSSITE:	PS50805; KRAAB; 1.
DR POSITE:	PS00028; ZINC_FINGER_C2H2_1; 19.
DR POSITE:	PS50157; ZINC_FINGER_C2H2_2; 22.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding; Nuclear protein; Repeat. 	
FT DOMAIN 1 69 KRAAB. FT 168 778 ZINC FINGERS. FT 168 190 C2H2-TYPE. FT 196 218 C2H2-TYPE (DEGENERATE). FT 224 246 C2H2-TYPE (DEGENERATE). FT 252 274 C2H2-TYPE (DEGENERATE). FT 280 302 C2H2-TYPE. FT 308 330 C2H2-TYPE. FT 336 358 C2H2-TYPE. FT 364 386 C2H2-TYPE. FT 392 414 C2H2-TYPE. FT 420 442 C2H2-TYPE. FT 448 470 C2H2-TYPE. FT 476 498 C2H2-TYPE. FT 504 526 C2H2-TYPE. FT 532 554 C2H2-TYPE. FT 560 582 C2H2-TYPE. FT 588 610 C2H2-TYPE. FT 616 638 C2H2-TYPE. FT 644 666 C2H2-TYPE. FT 672 694 C2H2-TYPE. FT 700 722 C2H2-TYPE. FT 728 750 C2H2-TYPE. FT 728 750 C2H2-TYPE.	

Best Local Similarity 23.98; Pred. No.6.le-05;
Matches 96; Conservatve 42; Mismatches 141; Indels 123; Gaps

351 SGEELGYGVETSSLEKPGGRHRCRCACAKVFGSDSALQILHSHTGERPYKCNCVGNRFTTR 410
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
290 SSRLIQHQRITHTGEP--YRCEECGKAFFGGSSLIITHRIHQIHGERPYCGREGKAFSQ 346
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
411 GNLKVFHRHREKRPRHYQMNPVRPHVDLVTSSSLPGLMGSMVRPEAKEAEAGTGGGVE 470
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
347 SOLV-----RHQTHTGPERPYCKE-----CGKAFFS 372
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
471 RKPLVASTATLSARESTLTSTGSTAFAVAPGLPTFKFVLMAKAVEPKSKADENTPPSEEG 530
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
373 QSSTLQHQRHNTGEKQILKASDS---PSLVAHQR---THAVEKPFKDEC-----G 419
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
531 SATAGVADSGSATMQL-----SKLVTSLEFWALLTNHLKS-TGSFPF----- 572
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
420 KARRWIS---RLSOHLTIHTGERPKYCKNKTCAFAFGCSRLIRHQRTHTGEPFYCDECGK 476
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
573 -----PYLEPLGASPSETSKL--ODLVKEIDRGAAVAVASTASGA 611
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
477 GPVQGSLLIQRHITHTGEXKPYVCNOCCGKAFFSSSSLIIHRIHNGEK----- 523
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
612 PTTSAAPASSASGPNOCVICLRVLSCPALRLHYQHGGERPPKCYCGRAFSTRGNLR 671
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
524 -----PYECLOGGKAFFSMSYQLTRHQHVHTGERPYKCNECGKAFFSONSTLF 569
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
672 AHFVGHKTPAAPAAQNASCPIOCKKFTFNATLOOHVMHNGGO 713
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
570 QHQTTH----AGVKRYECSECGKAFSRSSYLEHQRHITHTAQ 607
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 37
ID 2151 HUMAN STANDARD: PRT: 803 AA.
AC Q13105; O15932; Q9NUC9;
15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DD 15-JUN-2002 (Rel. 41, Last annotation update)
DE zinc finger protein 151 (Myc-interacting zinc finger protein) (Miz-1 protein).
GN ZNF151 OR MIZ1.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RA MEDLINE=97453565; PubMed=9308237;
RX Schneider A., Peukert K., Ellers M., Haemel F.;
RT "Association of Myc with the zinc-finger protein Miz-1 defines a novel pathway for gene regulation by Myc.";
RT Curr. Top. Microbiol. Immunol. 224:137-146(1997).
LN [2]
RA SEQUENCE FROM N.A.
RP Bird C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
LN [3]
RA SEQUENCE OF 580-803 FROM N.A.
RC TISSUE=Insulinoma;
RX MEDLINE=96044430; PubMed=7557990;
RA Tommerup N., Vissing H.;
RT "Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders.";
RT Genomics 27:264(1995).
LN [4]
RA SEQUENCE OF 327-342 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=92372070; PubMed=1505991;
RA Lichten P., Bray P., Ried T., David I.B., Ward D.C.;
RT "Clustering of C2-H2 zinc finger motif sequences within telomeric and fragile site regions of human chromosomes.";

BL	Genomics13:999-1007(1992).
-I-	SUBUNIT: ASSOCIATED WITH THE C-TERMINAL OF MYC.
CC	-I- SUBCELLULAR LOCATION: Nuclear (potential).
CC	-I- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
CC	-I- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; U20647; AAC50256.1; -
DR	EMBL; AL034555; CAB85445.1; -
DR	EMBL; Y09723; CAAT0889.1; -
DR	EMBL; M88369; AAA61327.1; -
DR	HSSP; P07248; IARE.
DR	TRANSFAC; T03414; -
DR	GeneW; HGNC:12936; ZNF151.
DR	MIM; 604084; -
DR	InterPro; IPRO00210; BTB_POZ.
DR	InterPro; IPRO00822; ZnF_C2H2.
DR	Pfam; PF00096; zf-C2H2; 13.
DR	Pfam; PF00651; BTB; 1.
DR	PRINTS; PR00048; ZINCFINGER.
DR	ProDom; PD000003; ZnF_C2H2; 2.
DR	SMART; SM00225; BTB; 1.
DR	SMART; SM00355; ZnF_C2H2; 13.
DR	PROSITE; PS50097; BTB; 1.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW	Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW	Nuclear protein; Repeat.
FT	DOMAIN 1 104 BTB.
FT	DOMAIN 306 739 ZINC_FINGERS.
FT	ZN_FING 306 739 C2H2-TYPE.
FT	ZN_FING 334 356 C2H2-TYPE.
FT	ZN_FING 362 384 C2H2-TYPE.
FT	ZN_FING 390 412 C2H2-TYPE.
FT	ZN_FING 418 440 C2H2-TYPE.
FT	ZN_FING 446 468 C2H2-TYPE.
FT	ZN_FING 474 496 C2H2-TYPE.
FT	ZN_FING 502 524 C2H2-TYPE.
FT	ZN_FING 530 552 C2H2-TYPE.
FT	ZN_FING 558 580 C2H2-TYPE.
FT	ZN_FING 586 608 C2H2-TYPE.
FT	ZN_FING 614 637 C2H2-TYPE.
FT	ZN_FING 717 739 C2H2-TYPE.
FT	CONFLICT 73 M->V (IN REF. 2).
SEQ	SEQUENCE 803 AA; 87959 MW; FEE72BA4E14AE200 CRC64;
QY	Query Match 5.0%; Score 263; DB 1; Length 803;
QY	Best Local Similarity 19.3%; Pred. No. 7.5e-05;
QY	Matches 178; Conservative 92; Mismatches 312; Indels 342; Gaps 31.
DG	73 SSASSAPRPECHSRSE-----OVMDTSHSNPPDSGGSGPPDPWTGPFR-RKEE 118
DG	113 SLAEPATSPGNAEAELATEGGCKRAKEEVAVSTLSRLDAQGRSTP---IGPSHDLEE 168
DG	119 SSGGFVAATGTAAAGGGGLIASPKLATPRPESTRPADPPPPP-----PPPPVGSGH 174
DG	169 RGGDAQSASGAEE-----QTEKADAFREPPVELKRPDPTSGMAAAE 209
DG	175 LNIPLI-----LEELRVLOQRDIHQOMTEIQCVLLIGSLGOTVGAPASESLRG 226
DG	210 AEALESSESQEWEVEYPAARKGEEOKEOEEOEEE-----GAGPAEYKE 252
DG	227 TGAASSTPLPLPFPIRKPAQTGKTAASSSSSSSSSGAAPPKCAFPHLHPHLSQHPSV 286
DG	253 EG-----SOLENGAPEENENEESACTDSGOE-----LGSE----- 283

287 GGVGRSHKPTAPSPALPGSTDLIASPHLAFPGTGTGLLAQCLAGRLAEMASPGLK 346
284 -----AKGLRSGTYGDRT 297
347 PRKNGSGELGYGEVVISLEKPRGHRKRCFAKVFSGSDALQIHLRSHTEGRRPYCNVGNR 406
298 SK-----AYGSVI-----HKCEDCKEFTHTGNFKRIIRIHTGKPKFCRCRCSKA 342
407 FTTGNGNLKVHFRHREKYPHYQMNPHVPEHLDYVITSSGLPYGMSVPEKAEKSTPG 466
343 FSDPRACSAHEKTH-----SPLKPYGC-----EECG-- 368
467 GGVGRKPLVASTALSAFESLTLSTGTSTAVALPGLPTPKFVLMKAVPEKSKADENTPP 526
369 -----KSYRLIS-----LNLHK-----K 382
527 GSEGSATAGVADSGSATRNQSLKVLVSLPFWALLTNHLKSTGSPFPVYLEPLGASPE- 585
383 RHSGEARVRCEDCG-----KLFTT--SGNLKRHLVHSGE--KPYCCDYCGRSGFSDP 430
586 TSKLOOL-VERIDRGAVAVASTAGAPTTASAPSSASGPNOCVILCLRVLSCPALRL 644
431 TSKRHLHTHDTKE-----HKCPHCDKKFNQVGNLKA 463
645 HYGONGGERPRKCKVCGRAFSTRGNLRAHFVGHKTSAPARAQNSCPICOKKFTNAVTLQ 704
464 HLKTHIDGPLKRCRCGKFTTSGNLKRHLRIH-----SGEKPYVICHORORADPALOR 519
705 HYRHLHG---QIPNGGALSCEGGAQENSSSEOSTASGSPSPPOSOQSPDEEMSEE 761
520 HVRHTEKRPQCVCVACGKAFQTQ-----ASSLIHVRHTEKRPVCEKRCGRK 566
762 EEEDEEEDVYDEDSLARGSESGEKAISVGDSEVSAAEEVATSVAPTYKEMD 821
567 FVQSSQLANLIRHNHNIRH-----KCSVCSKAFVNVGD 600
822 SNEKAPQHTLPPR-----PPDNT-DHPQPMQGTSDVSGAMEEAKLEGISSPM 871
601 LSKHIIHTGKRPYLCDCGRGFRNVDNLRSHTVHQKAGIKLEPEEGSESVVYTD 660
872 AALQEGEGSTPIVEELN-LPEAMKKDRESSGRKACEVCGOSPTOTALPEHQKTHK 930
661 DMTLATLAEALATVATQTLVVVGAATVADETVLKA-----EISKAVQVOVEDPN 712
931 DGPLFTCVFCRQGLDRATLKHKM 954
713 THIIYACDSCGDKFLDANSIAGHV 736
SUPL 38
008_HUMAN STANDARD: PRT: 1167 AA.
043345:
16-OCT-2001 (Rel. 40, Created)
15-OCT-2001 (Rel. 40, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Zinc finger protein 208.
ZNF208 OR ZNF91L.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheta; Primates; Catarrhini; Hominiidae; Homo.
NCBI_Taxid=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=98391764; Pubmed=9724325;
Eichler E.F., Hoffman S.M., Adamson A.A., Gordon L.A., McCreedy P.,
Lamerdin J.E., Mohrenweiser H.W.,
"Complex beta-satellite repeat structures and the expansion of the
zinc finger gene cluster in 19p12.";
Genome Res. 8:791-808(1998).
-!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
-!- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AC003973; AAB94784.1; -
DR HSSP: P25490; 1UBD.
DR Genew: HGNC:12999; ZNF208.
DR MIM: 606760; -
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 33.
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PR00046; ZINCFINGER.
DR ProDom: PD000003; Znf_C2H2; 17.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; Znf_C2H2; 33.
DR PROSITE: PS00805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 33.
DR PROSITE: PS5015; ZINC_FINGER_C2H2_2; 34.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 7 75
FT ZN_FING 201 1147
FT ZN_FING 201 223
FT ZN_FING 229 251
FT ZN_FING 257 279
FT ZN_FING 285 307
FT ZN_FING 313 335
FT ZN_FING 341 363
FT ZN_FING 369 391
FT ZN_FING 397 419
FT ZN_FING 425 447
FT ZN_FING 453 475
FT ZN_FING 481 503
FT ZN_FING 509 531
FT ZN_FING 537 559
FT ZN_FING 565 587
FT ZN_FING 593 615
FT ZN_FING 621 643
FT ZN_FING 649 671
FT ZN_FING 677 699
FT ZN_FING 705 727
FT ZN_FING 733 755
FT ZN_FING 761 783
FT ZN_FING 789 811
FT ZN_FING 817 839
FT ZN_FING 845 867
FT ZN_FING 873 895
FT ZN_FING 901 923
FT ZN_FING 929 951
FT ZN_FING 957 979
FT ZN_FING 985 1007
FT ZN_FING 1013 1035
FT ZN_FING 1041 1063
FT ZN_FING 1069 1091
FT ZN_FING 1097 1119
FT ZN_FING 1125 1147
SQ SEQUENCE 1167 AA; 134352 MW; E2184DF23B0D35E9 CRC64;
Query Match 5.0%; Score 263; DB 1; Length 1167;
Best Local Similarity 21.3%; Pred. Neg. 0.00011;
Matches 147; Conservative 84; Mismatches 284; Indels 176; Gaps 26;
Oy 358 EVISLEKPGRRKRCFAKVFSGSDALQIHLRSHTEGRRPYCNVGNRPTTRGNLKVHF 417

416 EYIHTGEKP---YKCECGKAFNWSNLMENKIHGTGEPYKCECGKGFWSSTLSYHK 472
418 HHRREYPRHYOMNHPREHLDVYTSSGLPGMSV-----PREKAEFGTP 465
473 KIHTEVKP-----YKCECGKAFNWSAILIKHKRIHTEGEPYKCEE---C 514
466 GGGVERKPLVASTTALSATE-----SLPFLSTGTST---AVAPGL-----PTF 505
515 GTFESKYSTLTTHKAIHAGEKPYKCEGKFTIKYSTLTTHKAIHAGEKPYKCEGKAF 574
506 NKFVLM---KAV---EPKSKADENTPPGSEGSALAGVADSGSATRMQLSKLYT----- 552
575 SKFSILTGHKVIHTEGEPYKCEE-----CGKAFNWSNLMENKIHTEGEPYK 623
553 -----SLPSWALLNHLKSTGSPFPFPPYVLEPLGASPSSETSKLO-----QVLEKIDRGAV 602
624 EECGKGFSTFSVLTGH-KVIHTEGEPYKCECGKAYKWSSTLSYHKIHTEGEPYKCEEC 682
603 AVASTASGAPTTAPAPSSASGPNOCVLCRLVLSCPRALRLHYGONGEPREPKCYGGR 662
683 GKAFNSASILIKHKRIHTEDEK---PYKCECGKFTFSKYSTLTTHKAIHAGEKPYKCEGK 740
663 ASTRONLNAHFNGHKTSAPARAONSCPTCOCKFTNAVTLQOHVNRHNLGGQIDNGSALS 722
741 AFSKSTILTKHKVIH---TGEPYKCECGKAYKWPSTLSYHKIHT-GEKPYKCECG 795
723 ECGGAOENSSFOSTASGSPFPQOSQOOPSPEEMSEDEDEEDVTDSDSLAGRG 782
796 KGFMSISILTKHEVINTG-----EKPYKCECG----- 823
783 SESGGEKA---ISVRGDSEVSGAE-----EVATSVAAPTTVKEMDSNEKAPDHTL 831
824 -----KAFSWSLVSFESHKTHAGEKPYKCEAGKAYMTFSLTKHKVIHTEGK----- 871
832 PPPPPRDNLHDHQPMPQ---GTSVYSGAMEEBAKLEGISSPMAALTOBEGSTPLVEEL 889
872 -----PYKCECGKAFNWSNLMENKIHTEGEPYKCE---EECD-----KAF 910
890 NLPKAMKPDGESSGGRK--ACEVCGSFPQTALAEHOKTHHKDGLFPCVPCROGFLDR 947
911 SMPSSLTEHKATHAGEKPYKCECGKAFSWSRLEHKATHAGEEP-YKCECGKAFNWS 969
948 ATLKKHMLLAHHOVPAPRAPHONITLSLV 978
970 SNLMEHKRIHTEGEPYKCECGKSFSTFSL 1000
RESULT 39
LAS_DROME
D GLAS_DROME STANDARD; PRT; 604 AA.
C P13360;
T 01-JAN-1990 (Rel. 13, Created)
T 01-JAN-1990 (Rel. 13, Last sequence update)
T 01-OCT-1996 (Rel. 34, Last annotation update)
E Glass protein.
N GL.
S Drosophila melanogaster (Fruit fly).
C Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
C Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
C Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
C NCBI_TaxID=7227;
N N
P SEQUENCE FROM N.A.
P MEDLINE=89365138; PubMed=2770860;
X Moses K., Ellis M.C., Rubin G.M.;
T "The glass gene encodes a zinc-finger protein required by Drosophila
T photoreceptor cells".
L Nature 340:531-536(1989).
C -!- FUNCTION: GLASS IS PROBABLY A TRANSCRIPTION FACTOR REQUIRED FOR
C GENE EXPRESSION SPECIFIC TO PHOTORECEPTOR CELLS.
C -!- SUBCELLULAR LOCATION: Nuclear (Potential).
C -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X15400; CA33450.1; -
DR PIR: S05447; S05447.
DR HSSP: P08046; 1A1I.
DR TRANSFAC: T00329; -
DR Flybase: FBgn0004618; Znf_C2H2.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 5.
DR PRINTS: PR00048; ZINCFINGER.
DR ProDom: PD000003; Znf_C2H2; 2.
DR SMART: SM00355; Znf_C2H2; 5.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 5.
DR Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Repeat; Vision; Nuclear protein.
KW DOMAIN 437 571 ZINC_FINGERS.
FT ZN_FING 437 459 C2H2-TYPE.
FT ZN_FING 465 487 C2H2-TYPE.
FT ZN_FING 493 515 C2H2-TYPE.
FT ZN_FING 521 543 C2H2-TYPE.
FT ZN_FING 549 571 C2H2-TYPE.
FT VARIANT 60 60 E -> O.
FT VARIANT 361 362 PM -> SL.
FT VARIANT 377 377 S -> T.
SQ SEQUENCE 604 AA; 62528 MM; 82B947C78366C071 CRC64;

Query Match 4.9%; Score 262; DB 1; Length 604;
Best Local Similarity 19.3%; Pred. No. 6.3e-05;
Matches 125; Conservative 48; Mismatches 171; Indels 304; Gaps 18;

QY 118 ESSGQFLVATNTAGGGGGLILSPKIGATPLPESTAPPPPPPPPP----- 168
DB 231 QSSGNNIGSNGSA---GVANASCYET--SAGTAAPPPPAAMYPSMSVNSMN 283
QY 169 ----GVSGHL-NIPILIEELRVLDQROIHOMQTEQICRVLILGSLGQVGAPASPS 222
DB 284 MTGNHGVGADAGVPM-----QCSGMNT-----PPSNS 313
QY 223 ELRGTGAASSTKRLPLPSPIKPAQTGKTATSSSSSSSSSGAEPKQAFNLYNPLGSOH 282
DB 314 TSAAMAAAVNVLYRPLSP-----GHVPASATYSFTADFRAPARTG-----LGALP 360
QY 283 PFSVGVGSRSHKPPAPSPAL-----PGSTDLIASPHIAFGTGLLAAQCICGAARGLE 337
DB 361 PMT---VGEKESPPSPANNSSLAGYPTGVGNGGYRPHKS-----PTSYAALALGLS 409
QY 338 AAASPLILKPKNSGELGVEVISLEKRGGRHK--GRCAKVFESDALQIHLRSHTG 394
DB 410 LSA---FDEDEDSNEDLDDEGSS---GGEMKPNLCRLCGKTKYARPTLKHILRHSG 461
QY 395 ERPKYKNCVGNRTTKNLKVHPIHREKRYHVQKNHPREHLDVYTSSGLPYGMSVP 454
DB 462 ERPRYKPDCKKSFSQAANLTAHVRTH----- 487
QY 455 PEKAEAEAGTPGGGVVERKPLVASTTALSATESLTLSTGTSTAVAVAPGLPTFNKFLMKAV 514
DB 488 ----- 487
QY 515 EPKSKADENTPPGSEGSALAGVADSGSATRMQLSKLYTSLPSWALLTNHLKSTGSPFPFY 574
DB 488 -----TGQKPF--- 493
QY 575 VLEPLGASPSSETSKLOQVLEKIDRGAVAVASTASGAPTTAPAPSSASGPNOCVYICLR 634
DB 494 -----RCPICDR 500

```

635 VLSCPALRLHYGHGCGRPFCKYCGRAFTSGNLRRAHFVGHKTSPPARAQNSCPICQK 634
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
501 RFSQSSSVTHMRTSHSGRPYCKSKSFSDSSTLTKHLRIH-----SGEKPYQCKLCLL 556
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
695 KFTNAVYTLQOHVHRMLHGOIPNGGSALSEGGAQAGENSFEQSTASPGP 742
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
557 RFSQSGNLNRMHRYHGNNSNGSNATGTVGG-----ESTSGSGVG 597

:SUPL 40
:93 MOUSE
:ZF93 MOUSE STANDARD: PRT: 645 AA.
061116;
16-OCT-2001 (Rel. 40, created)
16-OCT-2001 (Rel. 40, last sequence update)
16-OCT-2001 (Rel. 40, last annotation update)
zinc finger protein 93 (Zfp-93).
ZFP93.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
STRAIN=CD-1; TISSUE=Testis;
MEDLINE=96207307; PubMed=8617494;
Shannon M., Ashworth L.K., Mucenski M.L., Lamerdin J.E., Branscomb E.,
Stubs L.;
"Comparative analysis of a conserved zinc finger gene cluster on human
chromosome 19q and mouse chromosome 7.";
Genomics 33:112-120(1996).
-1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
-1- SUBCELLULAR LOCATION: Nuclear (Probable).
-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
-1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: U46186; AAB03529.1; -.
HSSP: P25490; 1ZNM.
MGD: MGI:107611; Zfp93.
InterPro: IPR001909; KRAB.
InterPro: IPR000822; Znf_C2H2.
Pfam: PF00096; zfc_C2H2; 13.
Pfam: PF01352; KRAB; 1.
PRINTS: PR00048; ZINC_FINGER.
PRODOM: PD000003; Znf_C2H2; 11.
SMART: SM00349; KRAB; 1.
SMART: SM00355; Znf_C2H2; 13.
PROSITE: PS50805; KRAB; 1.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 13.
PROSITE: PS50157; ZINC_FINGER_C2H2_2; 13.
Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
Nuclear protein; Repeat.
DOMAIN 8 86 KRAB.
ZINC_FINGERS.
ZN_FING 285 643 ZINC_FINGERS.
ZN_FING 285 643 ZINC_FINGERS.
ZN_FING 307 307 C2H2-TYPE.
ZN_FING 313 335 C2H2-TYPE.
ZN_FING 341 363 C2H2-TYPE.
ZN_FING 369 391 C2H2-TYPE.
ZN_FING 397 419 C2H2-TYPE.
ZN_FING 425 447 C2H2-TYPE.
ZN_FING 453 475 C2H2-TYPE.
ZN_FING 481 503 C2H2-TYPE.
ZN_FING 509 531 C2H2-TYPE.
ZN_FING 537 559 C2H2-TYPE.

```

```

FT ZN_FING 565 587 C2H2-TYPE.
FT ZN_FING 593 615 C2H2-TYPE.
FT ZN_FING 621 643 C2H2-TYPE.
SQ SEQUENCE 645 AA; 73012 MW; 7459CD14055AF469 CRC64;

Query Match 4.9%; Score 261.5; DB 1; Length 645;
Best Local Similarity 22.5%; Pred. No. 76-05;
Matches 126; Conservative 59; Mismatches 195; Indels 179; Gaps 21;

OY 180 ILEELRYLQGRQIHOMQTEQICRQVLLIGSGOTGAPASPSELPGTGAASRTPKPL 239
      | : | : | : | : | : | | | | | | | | | | | | | | | | | | | |
DB 211 VIFPMYLLTQHCVYREQAYQCSR-----GQEVPSDSPLEL-----HQOTILGK 255
OY 240 FSPYKPAQTGKTASSSSSSSSSGAEPPKQAFHLYHPILGSHQHPFSVGGRSHKPTAP 299
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 256 KSPYH--STHKDTHSPVPPIQPSVHPGRKRYW--CHEGKGF-----RQ 296
OY 300 SPALPGSTDLIASPHILAPGTTGLIAOCLGARGILEAASPGLPKNGSGELGYEV 359
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 297 SSAL--QTHQHY-----HTGEKPYKDCSGKGF-----SSSSDLNIHR 333
OY 360 ISSLEKPGGRHKRCRCARVFGSDSALOHLRSHTEGEPYCNVCNRPRTTGNLKVHFR 419
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 334 VHTGEKP---YKCEYCGKFTQMAHLQAHERRITGEKPYKCGDKRSCSNLTHQRY 390
OY 420 HREKYPHYQMNPHVPYEHLDVYITSSGLPYGMS-----VPEKAEERAGTPGG 467
      | : | : | : | : | : | | | | | | | | | | | | | | | | |
DB 391 HTEKRP-----YECNEGKRPSTLSGNDLHQHRYHTEKPYKCE-----CGK 432
OY 468 GVERKPLVASTTAUSATESLTLISTGTSTAVADGLPTFNKFLMKAVPKSADENTPPG 527
      | : | : | : | : | : | | | | | | | | | | | | | | | |
DB 433 GF-----SSASSFSQHVHTGEK-----PFH----- 454
OY 528 SEGSAIAGVADSGSATRMQLSKLVTS-----LPSMALITNHLKSTGSPFPYV 575
      | : | : | : | : | : | | | | | | | | | | | | | | | |
DB 455 ---CSVCKNFSRSSHFLDHQRHTHTEKPYRCVCGKRP--WLSLSLHSHQSVHTGKKPYK 510
      | : | : | : | : | : | | | | | | | | | | | | | | | |
OY 576 LEPUGASPSETSKLQOLYKIDRQGAVAVAASGAPTTAPADSSASGPNOCYICLVY 635
      | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 511 CGEGCKGFSHMSLQ-----AHHSVHTG-----EKPRCNVCORQ 545
OY 636 VLSCPALRLHYGHGCGRPFCKYCGRAFTSGNLRRAHFVGHKTSPPARAQNSCPIC 692
      | : | : | : | : | : | | | | | | | | | | | | | | | |
DB 546 FSKTSNLOAHORVHTHGEKPYKCDTCGKAFSOKSSLQYHQRHTHTEKPP-----KCEEC 598
OY 693 OKKFTNAVYTLQOHVHRMLHG 711
      | : | : | : | : | : | | | | | | | | | | | | | |
DB 599 GKFRWYVGLSHQRYHTG 617

```

Search completed: January 13, 2003, 15:19:06
Job time : 37.4828 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:17:45 : Search time 42.9357 Seconds
(without alignments) 4808.565 Million cell updates/sec

Title: US-09-988-117-3
Perfect score: 5301
Sequence: 1 MA0ETGSSSRLLGPGCEPAE.....SSIPSPGLSPPPRKDDPTMP 1002

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1265	23.9	1350	13	Q91929
2	1260.5	23.8	1377	13	Q9DDN5
3	1258.5	23.7	1261	13	Q802M5
4	1255.5	23.7	1308	13	Q9DF77
5	1165	22.0	1272	11	Q9EPW7
6	1142.5	21.6	1061	13	Q9PVN7
7	1055.5	19.9	826	13	Q9PVN6
8	1014.5	19.1	898	13	Q12958
9	1007.5	19.0	1053	4	Q9UJQ4
10	915.5	17.3	382	13	Q9W7G2
11	765	14.4	1373	5	Q9VKH2
12	742.5	14.0	1263	5	P91639
13	736.5	13.9	1267	5	Q9VKH3
14	663	12.5	329	13	Q9DGH3
15	654	12.3	327	13	Q9DGH1
16	627.5	11.8	288	13	Q9DGH2

17	526	9.9	549	13	Q9PVN5	Q9PVN5 xenopus lae
18	377.5	7.1	744	5	Q17396	Q17396 caenorhabd
19	349	6.6	2282	11	Q61479	Q61479 mus musculu
20	345	6.5	2406	4	Q9B2S0	Q9B2S0 homo sapien
21	337	6.4	927	4	Q9UEG4	Q9UEG4 homo sapien
22	336.5	6.3	1300	4	Q15090	Q15090 homo sapien
23	330.5	6.2	619	4	Q96K58	Q96K58 homo sapien
24	330	6.2	1173	13	Q90783	Q90783 homo sapien
25	325.5	6.1	1258	4	Q9P2A7	Q9P2A7 homo sapien
26	322.5	6.1	744	5	Q24219	Q24219 drosophila
27	321.5	6.1	498	4	Q9H8L4	Q9H8L4 homo sapien
28	320.5	6.0	751	4	Q9NS43	Q9NS43 homo sapien
29	316	6.0	1615	13	Q57415	Q57415 gallus galli
30	313.5	5.9	1237	11	Q9D2D7	Q9D2D7 mus musculu
31	312	5.9	734	4	Q96171	Q96171 homo sapien
32	311.5	5.9	744	5	Q9V0U9	Q9V0U9 drosophila
33	311.5	5.9	1920	5	Q46205	Q46205 drosophila
34	310	5.8	962	5	Q61360	Q61360 drosophila
35	309	5.8	736	5	Q960L6	Q960L6 drosophila
36	309	5.8	934	5	Q9YKFL	Q9YKFL drosophila
37	309	5.8	1893	5	Q9W4J1	Q9W4J1 drosophila
38	307.5	5.8	1891	5	Q77275	Q77275 drosophila
39	304	5.7	725	4	Q96S24	Q96S24 homo sapien
40	300.5	5.7	644	4	Q9NT61	Q9NT61 homo sapien
41	300.5	5.7	823	4	Q9H162	Q9H162 homo sapien
42	300.5	5.7	2232	5	P91365	P91365 caenorhabd
43	299	5.6	561	4	Q96BV0	Q96BV0 homo sapien
44	299	5.6	894	4	Q9C0K0	Q9C0K0 homo sapien
45	296.5	5.6	654	4	Q96CP9	Q96CP9 homo sapien
46	296.5	5.6	754	11	Q99J65	Q99J65 mus musculu
47	296	5.6	1104	4	Q9B243	Q9B243 homo sapien
48	295.5	5.6	563	11	Q921H7	Q921H7 mus musculu
49	295.5	5.6	591	11	Q8R0V0	Q8R0V0 mus musculu
50	294.5	5.6	700	11	Q99K53	Q99K53 mus musculu
51	294.5	5.6	701	11	Q9R164	Q9R164 mus musculu
52	294	5.5	835	4	Q9H165	Q9H165 mus musculu
53	292.5	5.5	812	11	Q99PV7	Q99PV7 mus musculu
54	292.5	5.5	995	11	Q35615	Q35615 mus musculu
55	292	5.5	624	11	Q62065	Q62065 mus musculu
56	292	5.5	726	11	Q924S6	Q924S6 mus musculu
57	291	5.5	783	11	Q9QY56	Q9QY56 mus musculu
58	290	5.5	1660	5	Q9NKN0	Q9NKN0 leishmania
59	289	5.5	814	11	Q70162	Q70162 mus musculu
60	289.5	5.5	784	4	Q00146	Q00146 homo sapien
61	288.5	5.4	884	11	Q99PV8	Q99PV8 mus musculu
62	287.5	5.4	662	5	Q9N360	Q9N360 caenorhabd
63	287	5.4	596	11	Q9CVF3	Q9CVF3 mus musculu
64	287	5.4	813	11	Q9JLF8	Q9JLF8 mus musculu
65	286.5	5.4	1173	11	Q63624	Q63624 rattus norv
66	286.5	5.4	1190	4	Q99621	Q99621 homo sapien
67	285.5	5.4	567	4	Q9BWE0	Q9BWE0 homo sapien
68	285.5	5.4	567	4	Q9BU26	Q9BU26 homo sapien
69	285	5.4	1606	11	Q924A2	Q924A2 mus musculu
70	283	5.3	1453	4	Q9Y6T1	Q9Y6T1 homo sapien
71	282.5	5.3	841	11	Q9N2H2	Q9N2H2 homo sapien
72	282.5	5.3	756	5	Q61362	Q61362 drosophila
73	281.5	5.3	1608	4	Q96RK0	Q96RK0 homo sapien
74	281	5.3	703	11	Q9R161	Q9R161 mus musculu
75	280	5.3	650	11	Q91WF9	Q91WF9 mus musculu
76	279	5.2	744	11	Q62788	Q62788 rattus norv
77	278	5.2	3942	11	Q88737	Q88737 mus musculu
78	278	5.2	568	11	Q8R0R2	Q8R0R2 mus musculu
79	277.5	5.2	599	4	Q9F2F9	Q9F2F9 homo sapien
80	277.5	5.2	2703	5	Q9VEG7	Q9VEG7 drosophila
81	277.5	5.2	540	11	Q9JIB9	Q9JIB9 mus musculu
82	277	5.2	618	4	Q96MX3	Q96MX3 homo sapien
83	277	5.2	2715	5	Q61603	Q61603 drosophila
84	275.5	5.2	977	4	Q96H84	Q96H84 homo sapien
85	275.5	5.2	581	11	Q9ERR8	Q9ERR8 mus musculu
86	275	5.2	743	11	Q8R5D1	Q8R5D1 mus musculu
87	275	5.2	1312	4	Q9NR59	Q9NR59 homo sapien
88	275	5.2	556	4	Q9UMP5	Q9UMP5 homo sapien
89	273.5	5.2				

Spatl transcription factor Sall1.
SALL1.

Xenopus laevis (African clawed frog).
Xenopus laevis (African clawed frog).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetozoa;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae,
Xenopodinae; Xenopus.
NCBI_TaxID=8335;

(1)
SEQUENCE FROM N.A.
Klingbeil P., Frazzetto G., Bouwmeester T.;
"xsal1", a xenopus homolog of the human Townes-Brooks syndrome gene
SALL1."

Submitted (Oct-2000) to the EMBL/GenBank/DDBJ databases.
EMBL: AF310007; AAG45108.1; -

HSSP: p15822; LBOO.

InterPro: IPR000345; Cyt_c_heme_bind.

Ileuko: IPRO00822; Znf.C2H2.

Pfam: PF00096; ZnF_C2H2_10.

Prodrom: PD000003; Znf_C2H2_1.

SMART: SM00355; Znf_C2H2_10.

PROSITE: PS00190; CYTOCHROME_C2; UNKNOWN_1.

PROSITE: PS00028; ZINC_FINGER_CS2_1; 8-

PROSITE: PS50157; ZINC_FINGER_CS2_2; 9.

DNA-binding; Metal-binding; Zinc-finger.

SEQUENCE 1377 AA: 148861 MW: 6849FD2B35A16D CRC64:

Qy	537	ADSGATWQSLKVTYSLPSMALLTNHLKSTGSPFPVYVLEPLGASPSFEKSLQOLVEXI	596
Db	702	TDSGIS--VMFNNLL--LP---LMSQFNA--KFPFGGLDVTYPA--SETSKLQOLVENI	750
Qy	597	DROGAVAVASTASGAPPTTSAPAPSSSAGSPNOCVICRLVYLSCPRALRLHYGHGGERPFK	656
Db	751	DKK-----SSDPNECVICHRYLSCOSALKMHYRTHTGERPFK	787
Qy	657	KVCGRAPSTRGNLRAHFNGHKTSPARAARONSCPTOKFTNNAVTLQOAHYRMHGGQIPN	716
Db	788	KVCGRAPFTTKNNLTHTYSHVRAMPPLRVQHSCTPCOKKFTNAVVLQOAHYRMHGGQIPN	847
Qy	717	GSALSEGGAGAOENSSEQSTASGPGSFPQPOSQOPSPEEEMSEEEDEEEDVDYD--	774
Db	848	TPVA-----ENRPD--SMGSDTGSF-----DEFTIDDLNFDSDENMEDCPDSS	888
Qy	775	-----EDSLAG-----	791
Db	889	VPDTPKSIDASODSLSSSPLPLEVSSITALENOMKLINAGLAEQLQASLKSANGSVEGD	948
Qy	792	SVRGDSSEVS--GAEEVAIVASVAPPTVKEKDSNEKRAPQHTLPPPPPPDMLDPQPREOG	850
Db	949	GMTINSSSSIGCCMESQASGAPASESTYSM-----HALSPFNSTIDYLSKSPNDEK-	999
Qy	851	TSQVSGAMEEKEKLEGIS--SPMAALTQEEGTSTFVELNLNPEAKMKPDGESSG----	904
Db	1000	-----LORAVSLDPTNGLSPPA-----NGALDLTSS--NIDKYKEEP---LGLVLP	1043
Qy	905	-----RKACEVCGQSFPPTOTALEEHOKTHPKDGPFLTVCFCROGFLDRAFLKKHML	956
Db	1044	FDRCKYKNTICDICGKTACQSGALDIHYKSHTKERP-FICTYCNRGFSYKGNLKHQHLT	1102
Qy	957	AHHQV-----PPFAHQPN--IATL-----SLVPG-----	980
Db	1103	--HOMDLPQSLEFPSSMTPNPTIPSAPSNPPLATIIKTEFNGFMHSSODIDKEQPTTNI	1160
Qy	981	-CSSSIFSPGLSP 992	
Db	1161	VSSGSLPSSATSP 1173	
RESULT 3			
0902W5			
AC	0902W5	PRELIMINARY; PRT; 1261 AA.	
DT	01-DEC-2001 (Tremblrel, 19, Created)		
DT	01-DEC-2001 (Tremblrel, 19, last sequence update)		
DT	01-JUN-2002 (Tremblrel, 21, last annotation update)		
DE	Spalt 2 protein.		
GN	SAL2.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=21184122; PubMed=11287198;		
RA	Farrell E.R., Tosh G., Church E., Munsterberg A.E.;		
RT	*Cloning and expression of CSAL2, a new member of the spalt gene		
RT	family in chick. "		
RL	Mech. Dev. 102:227-230(2001).		
DR	EMBL; AF304358; AAK38570.1; -		
DR	InterPro; IPR000345; CytC_heme_bind.		
DR	InterPro; IPR000822; ZnF_C2H2.		
DR	pfam; PF00096; zf-C2H2; 7.		
DR	PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.		
DR	PROSITE; PS00157; ZINC_FINGER_C2H2_1; UNKNOWN_7.		
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.		
DR	DNA-binding; zinc-finger.		
SW	SEQUENCE 1261 AA; 134317 MW; BB22E257BCAEB565 CRC64;		

Query Match 23.7%; Score 1258.5; DB 13; Length 1261;
Best Local Similarity 31.0%; Pred. No. 1.9e-66;
Matches 367; Conservative 141; Mismatches 341; Indels 333; Gaps 39;

16 GEPARAGDASE-----EHHQVCAKCAQAFSDPTFEFLAHQNSCTDPPVAVI-----I 64
16 GEGADGGGSGNSRSGSEETNVCCECAEFKWTDFLEHKKCTKPNPLVLIWEDEAAP 89
65 GGOENPSSNSASAPREGHRSOVMDEHSNPPDSSGSPDPPTGPERGSESSQFL 124
90 PAEEPEPEPASP---PSDAQESA--AEEGVQPENSSESSEKSTEKEEPEMEVETAEKS 144
125 VAACTGAAGGGGGLIASPKLAGATPLPEESTPAPPPPPP----- 165
145 FQNGSTN-----TATPLPOLPEPSPMTSYTMTNTVLTLLSTKVAQAQF 191
166 -----PPPGVSG--HNIPLLEELRVLQORIHOMQMTBOICROYLLG-----S 210
192 SOSARSAASASISSGTAVAIPMLLEQLMALQOQOIHQLEQIRSVQVAMNRQPLRA 251
211 LGQTVGAPAPSELPCTGAAS-----STKPLPLPSPIKPAOTGKTT----- 252
252 LNPVVAAGV-----TGQASNLQGFATSAIQLTAVLPALIMQAAGAPPAFDGSH 305
253 -----ASSSSSSSSGAEPKQAFPHLYHPLGSHPSVG----- 287
306 ISRPTSGASTPNISSGSSAPRESSAPCSSNAITSVTPVSVSNNTTISAQOPQNASTPPI 365
288 GVGRSHKPTPAPSPALPGSTDOJLASPHLAFPGTTGLLAOCLGARGLEAASPGLKX 347
366 GHGSLTITSNLDPNPLPOT-----SSNSVTFPNPLVSTIA---TANALDPLSA--LMKH 414
348 KNGS--GELGYEYVSILEKPRGHNKCFKCAKVFSGSDSAQIHLRSHTGPERPYKCNVGNR 406
415 RKGRPNVSVFEPKSSSEDPFEKHKCFKCAKVFSGSDSAQIHLRSHTGPERPYKCNVGNR 474
407 FTTGNLKVHFRHREKTYHVQMNPNHVPREHLVYITSSGLPYGMSVPEK-----AEE 461
475 FSTGNLKVHFRHREKTYHVQMNPNHVPREHLVYITSSGLPYGMSVPEK-----AEE 461
462 AGTP-----GGGVKPRPLVASTVLTALSTESLTLSS-----TGSTAVAPGLPTFNKF 508
535 PVLPTIPISIGLQLPPTIGVNSYGDSPSTTPMRSRSPORPSPASSECTISLSLNTSESG 594
509 VLKRAVERK-----SKADE--NTPPGSE-----GSAIAGVADSGSATRMLSKLVT 552
595 VPVASESPQVQSSSVTKAEPISLPASTRLGDHSLGCVSTASTSSITPIYTDSSVST 654
553 SLSPWALLTNHLKSTGSPFPYVLEPLGASPSETSKLQOLYEKIDRQCAVAVASTAGAP 612
655 SLSPWALLTNHLKSTGSPFPYVLEPLGASPSETSKLQOLYEKIDRQCAVAVASTAGAP 612
613 TTSAAPSSSSAGPNCVCLRVLCSPRALRLHYGOGHGEPEFKVCGRAFSTRGNLRA 672
700 -----MTDPNOCVICHVLSGOSALKMHRTHGEPFKKICGRATTKGNLKT 749
673 HFVGHKTSPPARAQNSCPIQCKKFTNAVTLQOHVRLHGLGQIPNGSALSSEGGAA--- 728
750 HFVGHKTSPPARAQNSCPIQCKKFTNAVTLQOHVRLHGLGQIPNGSALSSEGGAA--- 728
729 -----QENSESGS-----TASGSGSFQPOQOPS 753
808 LSTDEKNTIDTSLNDDIDDSNEDEPELKDMSADSAKPLIYSGSCBSPSPSVIISIA 867
754 PEHEM-----SEEEDEDEEEDVDEDSLAGRG---SESGEKAIVRGDS 797
868 LBNQMKMIDSVNCOOLSLKSTENGSGESDHLNDSSSANGDLESQASGASAMS----- 922
798 EEVSGAEEVATVAAPTIVKEND-----SNEKAPQH---TLPPPPPPDNLDH 843
923 -----ESSSSMQALSPVNSSESFRRKSPGLSNQDEPQEIOLKTEKPDSPPP----- 969
844 PQPMEGS-----TSVSG--AAEEAKLEGISSPMAL---TQEGEGISTPLVEELNLE 893

Db 970 --ATENGALDLTSTNPGRPVKEEA-----PFSLLFLNREKRPSTPSLVTSTAPT 1020
QY 894 AMKD-----PGSSGRKACEVCG 912
Db 1021 MIKMNHNSKRPISLGEVPSLPAIGVAPAPQIVWSPGTPMPLADPPRRTPROHNCQSCG 1080
QY 913 QSPFQTALEHOKTHPKDQPLFTVCFCROGFLDRATLKKN 954
Db 1081 KTFSSASALQIHERHTHGERP--FGCTTCGRFTTKGNLKVHM 1121

RESULT 4
09DEF77 PRELIMINARY; PRT: 1308 AA.
AC 09DEF77;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Spalt 1.
GN SAL1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20442162; PubMed=10985862;
RA Faircl E.R., Munsterberg A.E.;
RT "Chick spalt 1 gene expression is controlled by a combination of FGF
RT and Wnt signals in developing limb buds."
RL Dev. Biol. 225:447-458(2000).
DR EMBL: AF288697; AAC13011.1;
DR HSSP: p15822; 1BBO.
DR InterPro: IPR000345; CytoC_heme_bind.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF00096; zf_C2H2_10.
DR ProDom: PD000003; ZnF_C2H2_1.
DR SMART: SM00355; ZnF_C2H2_9.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 1308 AA; 139846 MW; 81810CC16336298E CRC64;

Query Match 23.7%; Score 1255.5; DB 13; Length 1308;
Best Local Similarity 30.1%; Pred. No. 3e-66;
Matches 359; Conservative 148; Mismatches 320; Indels 365; Gaps 42;

QY 19 AERGD-----ASEHHNPQVCAKCAQAFSDPTFEFLAHQNSCTDPPVAVIIGGOENP 70
Db 22 SQRNGDLEKGGGNRTTKKDAHVCGRCAEFELSDLOHKKCTKQNLVLI---NENP 78
QY 71 SNSASSAPREGHNS-----RSQVMDTEHSNPPDSSGS 103
Db 79 ASPPETPPRPSPNPDQMDYVNTDQVDCSDLSEKRLKDEKESMDVEASSINNSSS 138
QY 104 GPDPPTGPERGSESSQFLVAAGTLAG-----GGGLI----- 139
Db 139 -----SKSVNNTSISNSSTMGTSVTTSLPHIGDITLTIGNSFVINSVITE 185
QY 140 -LASPKGATPLPEESTPAPPPPPPPPPPPVGSGHNLPLLEELRVLQORIHOMQMT 198
Db 186 NLQSTKVAVAFQSGEAR-----CNGASNNKLAVPALMQGLALQOQOIHQLOLI 234
QY 199 EQICROYVLGSLQCTVGAPAPSELPCTGAASVPLPLPSPK---PAQTKTASS 255
Db 235 EQIRHOILLIAS--QNTDMPISSSSQGTILRAS--NPLSTLSHLSQLAIAAGLAQSLA 291
QY 256 SSSSSSSGAE--PPKQ-----AFHLYHPLGSHPSVGVGV 290
Db 292 SOSASISGVKQPLPQLPQSNPGSTLIPSSSGSSPNINILAAVTTTPSSSEKVASISGSO 351

[illegible]

```

OY 644 LHYGHHGERPERFKCKVCGRAFASTGNLRAHFVGAHKTSPARQNSCPICOKKTNATLQ 703
Db 554 MHYRTHTGGERPFCKICIGRAFASTKSNLTKHYGHVHRANPLKLOHSCPICOKKFTNAVLO 613
OY 704 QHVRHMLGGOIPNG-----GSLMBEGGAQDENSEOS-----TAGS 740
Db 614 QHHRMHMGKIPPTPVSEASDDIDSMDEKNGELNNSSTIDEDLIDIMEDDELAENASG 673
OY 741 PGSEFPQSQSPSEEMSEEE-----EEDEREEDVTDEDSL-----AGRSESGGEK 789
Db 674 -----SKPPPHSTRAESPAMQSTGTGQDQPVTLPLSALNLOKNSVKSSENSESLE 725
OY 790 AIVSRDSEEVSGAEAEVATVAAPTTVKEMDSNKAPOHTLPPPPRPD-NLDHPQME 848
Db 726 SDGLTNDSSVWDQDEYPTGKS---PT-----QSEARFESPTNSQSDNSKSPSPSY 773
OY 849 QGTSIDVSGAMEEAKLEGISSPMALITDQEGESTRPLVELNLPEA-----MKKDPG 900
Db 774 NGIDDDGMLSKDEHSONG-----SLNDDGDGA-----LDTNGGFARKIKEEPG 817

RESULT 8
ID 012958 PRELIMINARY; PRT; 898 AA.
AC 012958;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Zinc finger protein sai (fragment).
OS Oryzias latipes (Medaka fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthopterygii; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RS SEQUENCE FROM N.A.
RA Slick R., Koester R., Witbrodt J.;
RT "Hedgehog signaling activates spalt at the midbrain-hindbrain boundary
RT of fish.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U77376; AAB51127.1; -.
DR HSSP: P15822; IBBO.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF00096; zf_C2H2_7.
DR SMART: SM00355; ZnF_C2H2_7.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
KW DNA-binding; Metal-binding; Zinc-finger.
FT NON_TER 1
SQ SEQUENCE 898 AA; 96668 MW; 7BB1CF5D7A104287 CRC64;

Query Match 19.1%; Score 1014.5; DB 13; Length 898;
Best local similarity 32.8%; Pred. No. 3.7e-52;
Matches 266; Conservative 110; Mismatches 197; Indels 237; Gaps 29,

OY 327 ACGUGARGGAEAAAPGLKPKNGS-GEIGYGVISLEKPGGRHKCRFCAKVGSDSAL 385
Db 4 ASIAATNALDPLSA--LMKRRKGKRPNAVSVEDTKRSSDEPRFKHKRCAKVGSDSAL 61
OY 386 QIHILRSHTERPERYKCMNGNRRFTTNGNLVHFHRAHREKYPHYVOMNPPRVREHLDVYTSS 445
Db 62 QIHILRSHTERPERYKCMNGNRRFTTNGNLVHFHRAHREKYPHYVOMNPPRVREHLDVYTSS 121
OY 446 GLPYGMSVPEPK-----AEEAGTGGGGER 471
Db 122 GIPYGMSPLEPKPVTTWLDKRPVLEPTVPTSVALLDPLPPLSMIGFAESPSTLPTSRSPQ 181
OY 472 KPIVASTTALSATESILTLSTGTSTAVAP-----GLPTFNKFLVAKKAVPKSKADENT 524
Db 182 RHSPFSSDECASLSPNAVDTSTKTTTSPSPNPNLGSDGPRPLKFBGILLSPYSAARGENT 241

```

b	y	525	PGESEGA1-----AGVADSGSATRMQLSKVTSLPMA-----LILNHLKSTG	568
b	y	242	TITTTTQVLTLSTNTSTSSGSC---OVSSEISS-PSASNAHVPLPMISEQFRA--	295
b	y	569	SFPPEYVLEPLGASPSETSKLOQLVEKIDROGAVAVAASATSPASSASGPNO	628
b	y	296	KPFPGGLDSM--OTSSETSKLOQLVENIDRK-----MTDPNQ	330
b	y	629	CVCICRLVLSGRALRIHNGOGGEPPECKYCGRAPSTRGNLRHFVGKITSAAKONS	688
b	y	331	CVJCHRVLSCSALKMHRIHTGEPEPKCKIGRAFPTKGNIKTHGVHRSKPLRVQS	390
b	y	689	CPICQKKFTNAVTLQOHVRMLGQIPNGSALSSEGGAOENSSEOASTASGSPFPQ	748
b	y	391	CPICQKKFTNAVVLQOHIRHMGGQIPN--PLPE--SIQEMETDL-----SPDEKS	438
b	y	749	SQGPS-P-EEMSSEEEDEEEDDYTD-----	775
b	y	439	LDAWSNVDDLLDEMEQMDESDELKEBELDPSKYSPGSPTSMISSIAAMENOMKI	498
b	y	776	DSLAI-----RQSESGE-----KAISYRGSE-----EVSG-----	802
b	y	499	DSTANMTHSFQOKRAONSSSFGEADCTTSDLSAVGABEGSLGSPALSESQMOHS	558
b	y	803	-----AEEVATSVA-----PTTVKEMDSNEKARPHLTLPPLPPPNLDHPOME--	848
b	y	559	PAHSHSESORSKSPAALNNNSNTSMTVGEQENNTAGLTV-----KSEKETSPSLSAI	613
b	y	849	--OOSTSVSGA-----MEERAKL-----EGISSP-----MALTOGEG----	880
b	y	614	EGTGALDLITATOPSRHIYIKESHFSMLFLNRDRGLSAPFLASTASIMIMENHGKSKS	673
b	y	881	-----TSTPLVEELINLEPAKKDKDESSGRACEVCGSFPTOTALTEH	924
b	y	674	DNHHLGIQVPAAAPTTPTMSPSINPLAPPRTPRTRKQN-----CQCCKNFSSASAQLH	729
b	y	925	OXTHPKDGPLETCVFCRQGFIDLRAFLKKHM	954
b	y	730	ERTHTGERP-FACSIGGRAFTTKNLKVHM	758
<hr/>				
RESULT 9				
d	g	09U04	PRELIMINARY;	PRF; 1053 AA.
d	g	09U04		
d	g	01-MAY-2000 (TREMBLrel. 13, Created)		
d	g	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
d	g	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
d	g	DJI112F19.1 (novel protein similar to SALL1 (sal (Drosophila)-like 1) (LOC57167)).		
d	g	DJI112F19.1		
d	g	Homo sapiens (human).		
d	g	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
d	g	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
d	g	NCBI_TaxID=9606;		
d	g	[1]		
d	g	SEQUENCE FROM N.A.		
d	g	Babbage A:		
d	g	Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.		
d	g	EMBL; AL034420; CAB61485.1; -		
d	g	InterPro; IPR000822; znf_C2H2.		
d	g	Pfam; PF00096; zf-C2H2; 8.		
d	g	PRODOM; PD000003; znf.C2H2; 1.		
d	g	SMART; SM00355; znf.C2H2; 7.		
d	g	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.		
d	g	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.		
d	g	DNA-binding; Metal-binding; Zinc-finger.		
d	g	SEQUENCE 1053 AA; 11231 MW; 61D0D1F21CB2B337 CRC64;		

Query Match	19.0%;	Score 1007.5;	DB 4;	Length 1053;
Best Local Similarity	29.2%;	Pred. No. 1.2e-51;		
Matches 313; Conservative	125;	Mismatches 341;	Indels 293;	Gaps 41;

[illegible]

Db 884 ALQIHERHTHGEKP-FVNCNICGRAFTTKGNLKVH-----YTHGANN 924

[illegible]

OS Drosophila melanogaster (Fruit fly)
OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
OC Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:
OC Ephyridioidea: Drosophilidae: Drosophila.
OX NCBI_taxid=7227;
[1]
RP SEQUENCE FROM N.A.
RN STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RX RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butlis K.C., Busam D.A., Butler H., Cadelo E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigbam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Putl V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
RA Shie B.-C., Siden-Ilanos I., Simpson M., Skupski M., Smith T.,
RA Spleer E., Spirding A.C., Stapleton M., Strong R., Sun E.,
RA Svitskys R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh K.-F., Zaveri J.S., Zhao M., Zhu G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003632; AAF53097.1; -.
DR HSSP: P15822; 1BBO.
DR FlyBase: FBgn0004579; salm.
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 7.
DR SMART: SM00355; znf_C2H2; 7.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
KW DNA-binding; Metal-binding; zinc-finger.
KW KW
SQ SEQUENCE 1373 AA; 151014 MW; A6B857870F645ACC CRC64;

Query Match 14.4% Score 765; DB: 5; Length 1373;
Best local Similarity 22.98; Pred. No. 4e-37;
Matches 319; Conservative 151; Mismatches 446; Indels 476; Gaps 46;

3 QETGSSSRLLGPGCEPAREGGDASEHNPOVCARKCAQFSDPEFLAHONSCCTDPVMV 62
Db 50 KQIGSQOEEENGCGSPLLTATTATTAASRSPDE-----PEEQGPEDOSTSRSI---PQST 100
QY 63 IIGGQENPSSSSASSADRPFGHSRSQYMDTEHSN---PPDSGSSGPPDP----- 109
Db 101 PPHOLENDIKSEAKSELEPEYEDNNNRVAMTKPSSEEREPEENASQSMSSPVAEASAEAAAT 160
QY 110 -GPEKRG-----SSGQFLVAATGTAAAGGGGLT-----LASPKIGAT 148
Db 161 ERTPEKEKEDVDEVDKPEAPSSAVSTETVTLTGGAAPATYLEAIQNNMIAQAPAAK 220


```
147 STIFNVQRHLMQLQLIQLHLSQLKRAEAAALGRSHSDSEEEPEPEPKQPTNLGK 206
235 PLPLPSPKPAOTGKTTASSSSSSSGAEPKQAFFHLYHLSGHPSS-----VGG 288
207 EELEEQGESEDDOESREENSKTDKRGTEBRKA-----EEGYSMMCDISSSIASS 260
289 VGRSHKPTPAPS-----PALPGSDQLASPHLAFPGTGTGLLAOCGAAGLEMAASPG 343
261 ITTNHDPAPAPNPNCLEMQRTEEVLD-----ASQSIHAAMQOE----- 303
344 LLKPKNGSELGYGEVVISLEKPGG--RHKCRCAKVFSGDSALQIHLBSHTGERPYKC 400
304 -----YSEYASKAEQSRGEIKFKHRCKYCGKIFGSYALQIHLRSHTGERPVC 351
401 NVCGNFTTRGNLKVHHRHREKYPHVQMNPRPVENHLDYVITSSG-----LP 448
352 NVCGSKFTTKGNLKVHQRTRQTLRPPMLLPVAPN-----VGHSGGQYQGEQYPIRLP 406
449 YGMSV-----PEKAEAEAGTPGCGVERKPLVASTTALSATESLTLLSTGTS 495
407 FAPPAVPQGEQONOVEEPEIRQEIYPVQAEDLSKPKVKE--KSHSPVEVKTKPE 464
496 TAVAPGLPTFNKFLVLMKAVBRKSADEMTTPGSESGALAGVADSGSATRMQLS----- 548
465 VKTDALPSESEK-----PEKEISK-----PVYTSRRNGSVKRRKQTSVSPQOE 508
549 ----KLYTSLPSTALLNLKSTGSFPFPYVLEPGA-----SPSTSKIQ 591
509 DREDDVEHHLIAKLVARSSASRESQPAEYSLAQMERIIDKSMDLEIDKTSSTSKIQ 568
592 IVEKIDROGAVAAVASTAGAPTAPASSASAGPNOCVIGLRLVSCPRALRLHYGCHG 651
569 LYDNLIE-----NKLTIDNQCIQCQAVMSGRSSLOMIHIRTGTG 605
652 ERPFKCVGGRFSTRGNLRAHFVGHKTSPPARAQNSCPICQKFTNAVTLQOHVHML 710
606 ERPFKRICGRAPATKONLKAHMSIHKIKPWRSQFCKVCVCHOKFSNGIILQOHIRITM 665
711 ----GGQ-IPNGSALSJEGGAQENSS-----EOSTAGPGSEFP 745
666 DDGSGGQGAAPANPGEERKIGIEDONSNKSLCTSDTLDTSTTSDHSGQSSGCGDFD 725
746 Q-----POSQPSD-EEMESEDEDEDEEDVTDSDS-----LAGRSGE 784
726 EEMTMDSTDSRONSSAATATPHPLERERDREKERIRIPNDCSDERSHNPDLTGGSE 785
785 SGGEKRAISVAGDSE-----VSGAEE-- 806
786 SEMPRAMDLSPPSSNSGRIFATGLANGAAGGSGNGGLPMLGMPMPNLLMAAAREEMH 845
807 -----VATSVAA----- 813
846 ALGHAAKAPPLRPFGLFGMLHPPPNVNCILCFKMLPSTLAALESHQSEAKERATGHAQ 905
814 -----PTTVKEMDSNEKAPQHTLPPRP--PPDNLDPQ 845
906 RPOCSDAGSPYGAKLTLNPNLFAKKRPSSSSSGEKLPESSNPPEAPENPAPPIKEDP 965
846 PMEOGTSIDVSGAMEEKLBGSISSPMALT--QEGEGTSTPLVE-----ELNL 891
966 DQEQ-----LWVEEGASAGEGSGTGATSNYPQEGDAEQSLKMQLLAHNRFPASPLDF 1018
892 PPA-MKKDGESS-----GRKACEVCGOSPTQVTALEEHOKTHPKXGPTFTCYFCRQGF 944
1019 QOALMSAGAPRTSSLDPPVNNKHFCVHCVRNPNSSSSALQIHMRHTGTDP--FOCANVCQAF 1077
945 LDRATLKKHMLLAHNOVPPAPHPQNTATISLVGCSSSIPSPPGLS 991
1078 TTKGNLKVHM-GTHMMTNPTRSRGRBRMSLELPMRQPNSGQGHGSS 1123
9DGH3
```

```
ID O9DGH3 PRELIMINARY; PRT; 299 AA.
AC O9DGH3:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative spalt protein (Fragment).
GN SPAL.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Camp E.M., Iardelli M.T.;
RT "PCR amplification of spalt exonic sequences from various
RT vertebrates.";
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ293862; CAC05355.1; -.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR SMART; SM00355; ZnF_C2H2; 4.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
KW DNA-binding; Metal-binding; zinc-finger.
FT NON_TER 1
FT NON_TER 299
SQ SEQUENCE 299 AA; 33064 MW; A9D03F130DA5FAB5 CRC64;

Query Match 12.5%; Score 663; DB 13; Length 299;
Best Local Similarity 41.7%; Pred. No. 7, 1e-32;
Matches 144; Conservative 49; Mismatches 66; Indels 86; Gaps 11;

QY 403 CGNRFTTRGNLKVHHRHREKYPHVQMNPRPVENHLDYVITSSGLPYGMSVPEKAEEA 462
DB 1 CGNRFTTRGNLKVHHRHREKYPHVQMNPRPVENHLDYVITSSGLPYGMSVPEKAEEA 462
QY 463 GTPGGVERKPLVASTTALSATESLTLLSTGSTVAVAGLPFF-----KFLVLMKAVEPK 517
DB 59 ----SMLDSKPYVLTSLTS-----SVGMILPPT--ISLPPKIKENNSLAISPSHS 105
QY 518 SKADE--NTPPGSESGALAGVADSGSATRM-----QLSKLYTSL----- 554
DB 106 AKSDGSPADTPMKNTDS-----VLEEGESTLTPTSNGKAENNOSSLLMSMSAVEGTIE 161
QY 555 -----PSWALLNLHLKSTGSFPFPYVLEPLGASPSETSLLQQLVEKIDROGAV 602
DB 162 YTTNSPPMATNPMLPMLMSQFKA--KFFFGGLDPL--QGESETSKLOQIVENIDRK--- 214
QY 603 AVASTASGAPPTISAPAPSSASAGPNOCVIGLRLVSCPRALRLHYGCHGGERPKKVCGR 662
DB 215 -----VADPNECVICHRILSCQSLKMKHYRTHTGERPKKVCGR 254
QY 663 AFSTRGNLRAHFVGHKTSPPARAQNSCPICQKFTNAVTLQOHVR 707
DB 255 AFTTKGNLKVHMSVHRAMPRLVQHSQICQKFTNAVTLQOHIR 299

RESULT 15
O9DGH1 PRELIMINARY; PRT; 327 AA.
ID O9DGH1
AC O9DGH1:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative spalt protein (Fragment).
GN SPAS3.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
```


[illegible]

Query Match	Best Local Similarity	Score	DB 13:	Length	288:
Matches 142:	Conservative 39:	Mismatches 69:	Indels 91:	Gaps 9:	
DR SMART; SM00355; znf_C2H2; 4.					
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.					
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.					
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.					
DR DNA-binding; Metal-binding; Zinc-finger.					
FW NON_TER 1					
FT NON_TER 1					
SO SEQUENCE 288 AA; 32261 MW; A07A1F2BDEA6F42 CRC64;					
Query Match	11.8%;	Score 627.5;	DB 13:	Length 288;	
Best Local Similarity	41.6%;	Pred. No. 8.7e-30;			
Matches 142:	Conservative 39:	Mismatches 69:	Indels 91:	Gaps 9:	
OY 403 CGNRFTTRGNLKFHFHRRREKYPHYOMNPHRPEHLDDYITSSGLPYGMSVPEKAEFA 462					
DB 1 CGNRFTTRGNLKFHFHRRREKYPHYOMNPHRPEHLDDYITSSGLPYGMSVPEKAEFA 55					
OY 463 GTPGGVEKRPVLAATASTESITLLSTGTSTAVAPGLPFFNFKVLMKAVEPKSADE 522					
DB 56 -PLSMVDSKPLVGNISIGFMSSSL-----PGLPV----IIKKEEGSVITKP 98					
OY 523 NTPPGSE-----GSMLAGVADSGSATPMQISK- 549					
DB 99 HSEVSELTTRKINGHOGEGFVCSPIIISNEKFOENVORLSVSTLRSRSGEDIAINTS 158					
OY 550 ----LVTSLESMALLTJNHLKSTGSPFPYVLEPLGASPSETSKLOOLVEKIDROGAVAVA 605					
DB 159 VNTGLITLKSEQLEAFLL--LGSLLP-----NPLGA--SETSKLEQLEVDNKK----- 203					
OY 606 STASGAPPTTSAPAPSSASGPNQVCYICLRVLSCPRALRLHYGOGHGERPKCKYCAAFS 665					
DB 204 -----YTDNECGICHRVLSQSLRMHMFHTHTGERPKRCYCGRAFT 246					
OY 666 TRGNLRAHFVGHKTSPPARAONSCPTCKKFTFNNAVYLLQOHV 706					
DB 247 TKGNLKTHYSIHRSMPRLIQNSCPICQCKFTFNNAVYLLQOHV 287					
RESULT 17					
O9PVNS	PRELIMINARY;	PRF;	549 AA.		
ID O9PVNS					
AC O9PVNS;					
DT 01-MAY-2000 (TREMBLrel. 13, Created)					
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)					
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)					
DE Xsal-3', long form (Fragment).					
GN XSal-3'.					
OS Xenopus laevis (African clawed frog).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;					
OC Xenopodidae; Xenopus.					
OX NCBI_TaxID=8355;					
RN [1]					
RP SEQUENCE FROM N.A.					
RA MEDLINE=99458636; PubMed=10527856;					
RA Onuma Y., Nishitakemura R., Takahashi S., Yokota T., Asashima M.;					
RT "Molecular cloning of a novel Xenopus spalt gene (Xsal-3)." ;					
RL Biochem. Biophys. Res. Commun. 264:151-156(1999).					
DR EMBL; AB030828; BAA85902.1; -.					
DR HSSP; P15822; 1BBO.					
DR InterPro; IPR000822; znf_C2H2.					
DR Pfam; PF00096; zf-C2H2; 5.					
DR ProDom; PD000003; znf_C2H2; 1.					
DR SMART; PS00355; znf_C2H2; 5.					
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.					
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.					
KW DNA-binding; Metal-binding; Zinc-finger.					
FT NON_TER 1					
SO SEQUENCE 549 AA; 59348 MW; ECBBA78D0F088A46 CRC64;					
Query Match	9.9%;	Score 526;	DB 13:	Length 549;	
Best Local Similarity	30.6%;	Pred. No. 2e-23;			
Matches 140;	Conservative 60;	Mismatches 123;	Indels 134;	Gaps 15;	

Query Match	7.1%	Score 377.5	DB 5	Length 744			
584	SETSLDQLV	EKIDROGAVAAVASTAGAP	TTSA	PAPSSASGPNQVCLRVLSCEPRAIR 643			
5	SETTKLOLV	ENLDNG-----		SETNECLICHRVLSCSLSLK 41			
644	LHYOQGG	ERPPKCVCCGRAPSTRGNLR	APHGHTSP	ARAQNSCPICQKFTNAVTLO 703			
42	MHYRTH	GERPPKCKICGRAPSTKSNL	KTHYVHRAN	PTLKLQHS	CPICQKFTNAVTLO 101		
704	QVHRMHL	GGQIPNG-----	-GSAL	SEGGGAQENSSES-----	TASG 740		
102	QHIRHMG	KGIPINTVYSEASDDIDSM	DEKNGELN	STDENLDDIM	DEDELAENASG 161		
741	PGSPFO	POSOQSPPEEKSEEE-----	EEDEEED	VDYDESL-----	AGRESGEEK 789		
162	-----	SKPTPTSE	TRAESPA	MOFTSTGTQDPV	PLPALNORONS	VKSSENGSLE 213	
790	ALSVRG	DEEYSGAAEEVATVAAP	-----	TVKEM	DSEKRAPOHTLPP	PPPPPN 840	
214	SDGLTND	SSVWDQETPTGKS---	PTQSE	ARFTSP	INSOSDSNA-----	KSPSYNG 263	
841	LD---	HPQ-----	PMEQ	TS	SDVS-----	GAM 858	
264	LDQGLM	SKDEHSQSLNPDGDAL	DLTNG	GFARKIKE	PEGLHONG	EGRGLNLYGAP 323	
859	EEBAKLE	ISSPMALTOE-GRG	STPLVEELN	PEAMK	KDGE	SSGRACEVQGSFPT 917	
324	PALIKEM	VS	DRMAATYLGPPN	IS	PG	INPLIVPOR-----RSAKO	HICTMGCRNFSS 377
918	QTAL	EHQTHPKDPLFTVC	RCRGGF	DRATL	KKHM 954		
378	ASALQI	HERHTG	KEP-FACT	ICGAF	TTKGLN	KYHV 413	

	Best Local Similarity	22.1%	Pred. No. 1.9e-14:	
	Matches	182;	Conservative	92; Mismatches 280; Indels 271; Gaps 29;
OY	22	GGDASEEHHPVCAKCAQESDPTEFLAHQNSCTDPPVMVIIIGQENBPSNASSAPRP	81	
		:::		
Dd	51	GGALPLEDRSNILPHFVFPANPOOFL-----SLCAQ-----LGNSSSRNVSTAS----	96	
OY	82	EGRSRSQVMOTESHNPDSGS---SGRDPDWGRRRGESSGOFLVAATGTAAAGGGGL	138	
Dd	97	-----TTSCPLQSCQSQSFSSPALIWHVLDAHEDEQIFESCDCYTTFISNQDI	146	
OY	139	-----IIASPRLGATPLPEESTPAPPPEPPPPPGVGSGHLNTPLLELRVLOORQ	191	
		:::		
Dd	147	REHKCKQTLLA-----RSTSVPESTIPSSVCFLSTPTTPLL-QEININSIGSEIREDEEE	202	
OY	192	IHHOMTEQLCRQVLLIGSLGQTVGARSPSELPTGTGAASSTRKPLPLDSPIKPAQTGKT	251	
		:::		
Dd	203	DMDVEGEHNAHQ--LFGHLIQ-----	224	
OY	252	TASSSSSSSSGGAEPPIQAEPFLHYHPLIGSQHPFSVGWGRSHKPRPASPALPGSTPOLI	311	
		:::		
Dd	225	DDSKMASLENHNAPPPAAPFPMPMPFLMRQPD-----PRADVEA	265	
OY	312	ASPHLAPPGTTGLIAAOCLGARGLEAAPSOLLKPKNSSGELGYGEVISLEKPCGRHK	371	
		:::		
Dd	266	AGRHDNDDDWEALMEISTSEAEKIRALV-----GDKAVPTTDP--NQ	306	
OY	372	CRCACAYEGSDSLQILHRSHTERPYKCYNCGNRFTRGNLKVHRHREKYPHYQMN	431	
		::: ::: :::		
Dd	307	CLICRYLSCKSALSQMRYRTHTGERPFCKLICQRAFTTKGNLKHMGVHRSH-----	359	
OY	432	HVPYELHDIVYTSSGLPYGMSVPEKA-----EEEAGTPGGGV--	469	
		: :::		
Dd	360	-----SERGLP--ISLPOLAMHQHOQIAPQRRIHNPPTSASAAMAAVAQ	406	
OY	470	-----ERKRPIVSTAALATE---SILT---LLSTGSTANA--PGLPTF-----	505	
		:::		
Dd	407	IQASQOCPLCQQRLNAGELAVHTEHRNSTLQPRVMTPTTTRQTPPEPVFFTPPS	466	
OY	506	-----NKFEVLKAVEPKSRKADENTPPGSGSAIAGVADSASATRMQLSKLTSPSWA	558	
		:::		
Dd	467	LNAIDMSTQNLNIISAOLKNDSS--PNTDTSSV-----EEKITRDDPPKMASLSPS--	517	
OY	559	LITNLHKSTGSEFPFYLPLEPLGASPSSETSILQOLEVK--IDROGAVA-----VAST	607	
		:::		
Dd	518	----NSSDSSS-----VRODILESSSEEEKKLKEPPILEQOVSYTPPNKENPPLAMQ	569	
OY	608	ASGAPITTSAPAFSSASAGPPNCVICRLRVLCSPRALRLHYHGOGGEPFCFKCGRAFSTR	667	
		:::		
Dd	570	KMAAETLEPPRPROMPVLSKHQCGVCFKHFSSSSALOIMHRTHTGDKPFCDMCGRAFFTTR	629	
OY	668	GNI RAHFVGH--KTSAPAARA-----	685	
		:::		
Dd	630	GNLKVMMGTSMQCSRSRRGRIRFDVASSVYEKPMILQSPIPTSGAPGASPLAMLGPNGL	689	
OY	686	-----QNSCPICQKKFTNAVTLQOHVBMHLGGQIPNGSA	720	
		::: :::		
Dd	690	SGLMEMMLMRTVCSVCQKVCQSPNELEQHLEKHLN-----NGSSA	730	
		:::		
RESULT	19			
ID	061479			
AC	061479:	PRELIMINARY;	PRT;	2282 AA.
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DE	01-MAR-2002	(TREMBLrel. 20, Last annotation update)		
Dd		DNA binding protein Rc.		
GN		KRC OR RC.		
OS		Mus musculus (Mouse).		
CC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
NC		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX		NCBI_TaxId=10090;		

RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
 RA Tanaka T., Kotani K., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AB002324; BAA20784.1; -.
 DR HSPF; P08047; I5P2.
 DR InterPro; IPR000822; ZnF_C2H2.
 DR Pfam; PF00096; zf-C2H2; 19.
 DR PRINTS; PR00048; ZINCINGER.
 DR SMART; SM00355; znf_C2H2; 19.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 19.
 KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
 KW NON_TER
 SQ SEQUENCE 927 AA: 102756 MW; 176C78A9CA07F8C0 CRC64;

Matches 227; Conservative 92; Mismatches 402; Indels 350; Gaps 45;

QY 67 QENPSSASSAPR-----PEG-HSRQVMDTEHSNPPDSG--SSG--PPDPTWGE 113

Db 10 EKDPRLGAOSVPRGRALKGLSPGLDSAFRLFDPDRAGPWNTAVI,SSGMFETAIWGPD 69

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

QY 114 RGESSGQFLVAATGTAA-----GGGGILASP-----KLGATPLPPESTPAP 158

```

Db  70 LQPEQSPN--DAHRGAESNEEESPRQESSGEEIIMGDPAQSPESKDSSTMSLERSSOD 127

```

On **150** PRRRPPRRPVCVCCCUU N T D I T I E R D U ODDOTHOVOEMO.TDPOW : : QARCCSWAAT CIO

2Y 159 PPPPPPPPGVSGHLNIPLLLEELRVLQQRQIHQMOMTEQICRQVLLGSLGQTVGAP 218

Db 128 P S V P Q N P P T P --- L G H S N - P L --- D H Q I P L D --- P 152

0x 219 ASPSEL.P-----CTGAASSTKBILPIESD-TRDQOTCXTMDeeeeeeeeee 363

Db 153 PAPERVPTPSDWTKACEASWQMGALTTWNSPPVVPANEPSSLRELVOGR----- 200

QY 263 GAEPKQAEFHL YHPLGSOHPFESVGVGRSHKPTPAPSPALPGSTDOLI ASPHIAFP GTT 322

[illegible]

Db 201 -----PAGAEKPYICNECKS-----FSQWS 221

QY 323 GLLAQCLGARGLEAASPGLLKPKNGSGELGYGEVISSLEKPGGRHKCRFCACKVFCS 382

[illegible]

DD 222 KLLRHQRIHTGERPNICSECG--KSFYQSSHLVQHQRTHTGEKP---YKCPDCGKCFSSWS 276

QY 383 SALQIHLRSHTERPYKCNCVCGNRETTGRLKVHFHRHREKYP----- 425

Db 277 SNLYOHTHTGEEKPYKCTECEKAFTOSTNI.TKHORSHTGEEKPYKCGECBPBAFVBSDDIT 336

277 UNDEQUANTHIGENT INCI ECENAF IQSINLIHQKSHIGENFIACGCEKKAFFIKSSDLI 336

QY 426 -----HVQMPHPVPEHLDYVITSSGLPYGMSVPPEKAE--EAGTPGGGVERKPLVASTT 479

Db 337 QHQAHTGEEKPYKCP-----CGKRFQGNHNLKHKIHA-----EKPYRCTEC 381

Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100																																																																																																																																																																																																																																																																																												
480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847

QY 480 ALSATESLTLTS-TGTSTAVAP-----GLPTFNKFVLMKAVEPKSKADENTPPGSESGA 532

Db 382 GKSFIQSSSELTQHQRTHTGEKPYECLECGKSGHSSTLIKHQRTHLREDPFKCP----- 435

0V 533 TAGVADSGSATBMOISKIYTSI.PSWAITNHUKSTCSEBEPVYI EBI CAEPCEMXYI --C EOO

```

000 1A0VAD0GSAIKMQLSKLV1SLPSWALLINHLKSLGSEFPYVLEPLGASPSSEISKL--Q 590
      : | : ||| | | : | | | | | : | |
001

```

Db 436 VCGKTFLLSAT-----LLRHQRTHTGE--RPHYKCEGCKSFSSVSSNLINH 478

0Y 591 OLVEKIDROGAVAVASTASCAPTTSAPAPSSSASGPNOCVTCI.RVI.\$CBRAI.RI.HYGONG 650

[illegible]

Db 479 QRIHGER-----PYICADCGKSFIMSSTLIRHQRIHT 511

QY 651 GERPFCKVCGRASTRGNLRAHEVGHKTSAPARAONSCPICOKETNAVTLLOOHVRMHL 710

[illegible]

Db 512 GEKPYKCSDCGKSFIR-----SSHLIQHRRHTGEEKPYKCPECCKSFSQSSNLITHVRTHM 567

QY 711 GGQI---PNGSALSEGGAQENSSEQSTASGPGSFQPQPQSQQPSPPEEMSEEEEEDEE 767

568 DENLFVCSDCSKAFLEAHNELJEOHRIHER-----GKTPARRAQ----- 605
768 EEEEDVTEDESLAGRSES-----GGEK-----AISVRGDSSEVSGAEEEVATSVAAPTTVK 818
606 -----GDSLTGLGIDPRLSLTPRRPGAKPHKGLVCGKGFNDE--GIFMORITHIGENPYK 656
819 EMDSENAKAPQHLTPRRPPRDNLHNPOMEQGSTDVSGAMEEBAKL-----EG1 867
657 NADG---LIINAAPKRP-----QLRSRPLRRGNS--YFGAAGRAAPROPPLKRPPEGEGF 708
868 SSPMAITQF-----GE-----GSTPLV---BELNLPEAMKKDPGES 902
709 SORGLSLSKTYICSHCESFLDRSVLHQHLTHGNKEKFLFPDYRIGLEGAGSPPLS 768
903 SGRKACEVCGSFPQTALAEHQKH-----PKDG-----PL---FTC 937
769 GKRPKCEPKQSGFLSSPELLHOKYHAGKSSOKSPELGSSVLEHLRLSPDGARPYRC 828
938 VFCRQGLDRATLTKKMLLHNVPPFAPHPGPNITATLSLPGCSSITSP 988
829 SDCRASFLDRVALTRHGOETHNQEKPRNPEDPRPEAVTLSTQEGEGETPTP 879

RESULT 22

PRELIMINARY: PRT: 1300 AA.

015090
01-JAN-1998 (TRMBLrel. 05, Created)
01-JAN-1998 (TRMBLrel. 05, Last sequence update)
01-JUN-2002 (TRMBLrel. 21, Last annotation update)
KIAA0390 protein.
KIAA0390.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_TaxID=9606;
SEQUENCE FROM N.A.
TISSUE-BRAIN.
MEDLINE=97349984; PubMed=9205841;
Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
Tanaka A., Kotani H., Nomura N., Ohara O.;
Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.;
DNA Res. 4:141-150(1997).
-1 SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
EMBL: AB002388; BAA20844.1; -.
HSSP: P15822; 1BBO.
InterPro: IPR000822; Znf_C2H2.
Pfam: PF00096; zf-C2H2; 10.
PRINTS: PR00048; ZINC_FINGER.
PRODOM: PD000003; Znf_C2H2; 1.
SMART: SM00355; Znf_C2H2; 10.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 6.
PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
DNA-binding: Metal-binding; Nuclear protein; zinc-finger.
SEQUENCE 1300 AA: 14146 MW: 85855B7047653664 CRC64;

Query Match 6.3%; Score 336.5; DB 4; Length 1300;
Best local similarity 20.8%; Pred. No. 1e-11;
Matches 193; Conservative 102; Mismatches 317; Indels 317; Gaps 37;

258 SSSSSGAEPKQAFPHLKHLP-LGSHPFVSGVGRSKH-----PTPAPSPAL 303
11 SSAFEKPEP-----HLSGPVLNGQVAMS---QKLHQITSQLSHAPPELHPRPNEEKRP 60
304 PGSTDOIASPHLAFP-GTTGLIAOCLGARGLEAASPLLPKNGSGELGCVGEVVIS 362
61 PASLEEKRAHYMSGOPMGSOMALLANQLEGREVDTSLNGRVDLQGLNQ-NIGINSQMSD 119
363 LE---KPGGRHKRCARCAVFGSDSALQILHLSHTGERPPYKCNVCNRPETTRGNLKVHFR 419

Db 120 IEDDARKNRKTPRCPCGKRRFRNSILSHMRTHTEGEPKPCPYCDHRAOKGNLKHILT 179
Qy 420 H-----REKYPVONPHVPEHL---DYVITSSGL-PYGSVPEPKAEAGTPEG 467
Db 180 HKLGNLGKGRVREBNRLHLEERAILRDKQLKSLDRPDLKPRPHNAQ----- 232
Qy 468 GVEKPLVASTALSTESLTLLSTSTAVAPGLPFPFNKFLKAVLPKSKADENTPRG 527
Db 233 ---APLAAGTALQANHSV-----PDVAHPVPS-----PK-----PAS 262
Qy 528 SEGSAIGVAD-----GSATRMQLSKVLTSLPSMALTLNHLKSTGSFPFVYLEPLGA 581
Db 263 VQEDAVAPAAFRCTGCKFKFKRE-----ELDRHITLHK-PYKSTICLDFPA 309
Qy 582 SPSETSKLQOLVEKIDROGAVAVASTAGAPPTSAPAPSSASGPNQCVICLRLVSCPRA 641
Db 310 SOEE-----ELISHVEK-----AHITSAOGGQPN----- 335
Qy 642 LRLHYGGHGER---PFCKYCGRAFTSGNLRAHPYGHKTSAPARAQNSCPICQKKFTN 698
Db 336 -----GGGQSANEFCEYCGVFSQAMFLKGHMKRHKDS---FHHCOICGRFKE 384
Qy 699 AVTLQOHVRMHLG-----GQIPNGSALSSEGGGAOE----- 730
Db 385 PWFLKNMKVHLNKLSTYKNSKSPSPDEVPYPMGMSQZAHANLYSRYLSCLOSGETPRKA 444
Qy 731 NSSEOSTASGSPFPQSOQ---PSPEEWSSEEEDEE----- 767
Db 445 GLSEPSOLYGKGLPMKEKELGLKLTSPISMAHGVEGDKHSLGLNLVPLKSSCIE 504
Qy 768 -----EEDVDYD-----EDSLAGSSESGEKA 790
Db 505 RLQAAKAAEKDPVNSYQAWOLMARGAMEHGLSKENHPLQRNHEDTLNAGVLPDKER 564
Qy 791 --ISVRGD-SEEVSGAEEVATSVAA---PTTVKEMDS-----NEKAQ 828
Db 565 EYLVGADGSKQKMPADLVHSTKVSQRLPSKLDPLESSDPLSHGILNQTLEYNLQGG 624
Qy 829 HTLPPEPPPP-----NID-----HPQWE 848
Db 625 NMKEKPTCEPCPGRYFRTYHQVHVSRVHKRDRKKEEDGLHVGDERRSGSDQESQSVS 684
Qy 849 QGTSVSGAMEEAKLEGISSPMALTOGEGSTPLVEELN-----LPEAMKKDP 899
Db 685 KSTTPGSSNVTSESGVGGSLGTSQAQEDSPHPSSESSSDIGEBAGRSAGVOOPLLR 744
Qy 900 GESSGRACEVCGSFPQTALAEHQKTHPKDGPLFTCFRCRQGLDRATLTKKMLLHNV 959
Db 745 SLGSAMKDCRYCGKTFRSHNLKVHRLRHTGKRP-YKCPHCDVAGTQSASLKYH-LERHN 802
Qy 960 Q-----VPPFADHPQONITLSL 977
Db 803 RERONGAGPLSGOPPNODHKDEMSSKASL 831

RESULT 23

PRELIMINARY: PRT: 619 AA.

Q96K58
AC Q96K58;
DT 01-DEC-2001 (TRMBLrel. 19, Created)
DT 01-DEC-2001 (TRMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRMBLrel. 21, Last annotation update)
DE CDNA FLJ14492 fls, clone MAMMA1002937, weakly similar to zinc finger protein 135.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY GLAND;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

Db 381 -----SHLOYHLRLHTGERPYRCPDCP-----KA 404
QY 521 DENTPPSEGSALAGVADSGSATFRMOLSKLVTSLPsmallTLNLKSTGSPFPVYLEPLG 580
Db 405 FKNT-----SCLGPH-----414
QY 581 ASPBETSKLOOLVEKIDROGNAVAVASTAGAPTTNAPAPSSASAGNOCVICTRLVSCPR 640
Db 415 -----RDL-----HTGERPHACPTICGKAFOTGS 437
QY 641 ALRLHYGNGGERPCKVCGRAFSTRGNLRAHFVGHKTSAPAARONSCPTCOKKFTNAV 700
Db 438 NLROHORHTGERPACSHCGKTFTHSSNQLH---QRTSSAR-PHCPCPLCPKAFVMS 493
QY 701 TLQOHVNMHLGQIPIPGSALSSEGGAAOENSSBOSTAGSGSPPOSOQSPREEMSE 760
Db 494 YLQRLRLFTN-----AAGPKGSRP-----512
QY 761 EEEDEEEEDVTDSDSLAGRSESGEKAISVNGDEEVSGAEDEVATSVAPTTVKEM 820
Db 513 -----ALTPORDGPVLOAA---LSLEVTAPE-----534
QY 821 DSNF-----KAPO-HTLPPPPPPDNL-----DHQPMEGTSDVSGAMEEAKLEGIS 868
Db 535 DAHPEFLLOTPQGLQIPSPBPAPQKILLLPTAPQPPKHGSGSPTPGOSLLLPSTGTT 594
QY 869 SP---MAALTOGEGCTSPVLEELNLPKAMKKDRGESSGRACE 909
Db 595 LPTURLQAVTAVPOGTGTGLVLQGLPEORPLHPAGIPRGOAAVE 638
RESULT 25
Q9P2A7 ID Q9P2A7 PRELIMINARY; PRT; 1258 AA.
AC Q9P2A7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE KIAA1441 protein (Fragment).
GN KIAA1441.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
DR EMBL; AB037862; BAA92679.1; .
DR HSSP; P17789; ZDRP.
DR InterPro; IPR000345; Cytochrome_bld.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 12.
DR SMART; SMD0355; Znf_C2H2; 12.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00128; ZINC_FINGER_C2H2_1; 9.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
FT DNA-binding; Metal-binding; zinc-finger.
FT NON_TER 1
SO SEQUENCE 1258 AA; 131684 MW; 04D3F1BD497782 CRC64;
Query Match 6.1%; Score 325.5; DB 4; Length 1258;
Best Local Similarity 20.4%; Pred. No. 4.3e-11;
Matches 245; Conservative 126; Mismatches 445; Indels 385; Gaps 54;
QY 12 GCGCGEPAERGGDASE-----HHPOVCAKCAOFSDPTEF-----LAHQNSCCTDPPV 60
Db 59 GGP-GRPEPGVSESEBDTAAASAGDGPVPAQASDHGLPPRDISVVSIVKNTVCPQOSE 117

QY 61 MYIIGQENPSSASAPRPEGHSRQ-----VMDTEHNPDPGSSGPPDPWG 111
Db 118 ALAGSAGDGAQADAVTKGGRVPHRMONGFSGPEPLPGTRPHSPAPRSG-----TW- 170
QY 112 PERGEESGQFLVAATGTAAAGGGGLILASPKLGAPLP-----PE-----STAPPP 161
Db 171 -KEKME-----GKPRLDLFANHGPERGDHSDPLPSSA 202
QY 162 PPP-----PPP-----PGVSGHLNIPLLLEELRVLOQOI--HQMOTGQ 200
Db 203 PSPREGALTPPFPFSSFELAENGPG-----MOPVSSPPLGALKQESCSPHNPQVLAQ 257
QY 201 ICRQVLLSLGQTVAGAPSPSELPGTA-----ASTKPLDLPFSPIKPAQOTGK 250
Db 258 QGS-----GSSPATDIPASAPPPVAGVFFFKOSPGHQSPLASKVPVCOPLKEEDDE 312
QY 251 TTAASSS-----SSSSSGAE-----PRKQAFHLVPLGQHPFSV-----GGV 289
Db 313 GPVDRSSPSPGSPSSGAEAADEDSNDSPASS-----SSRPLKVIKTIKTSQNI 363
QY 290 GR-----SHKPPAP-----SPALPGSTDQLIASPHL-----316
Db 364 TRTVTVQVPSPDPAPPLAEGAFILAASILKLSPATPTSEGPKVVSVLQDGTRLKGTLP 423
QY 317 -----AFPGTTGL-----LAAOCLGARGLEAASPG 343
Db 424 VATIQNASTAMLAASVARKAAVPLRPGGATSPKMLAKNVGLVPAQLPRADGAGLGTTG 483
QY 344 -----LKR-KNGSG-ELGYEYIS-----361
Db 484 OKVNGASVWVQPSKATPTSGTGTGTVISRTOSLVEAFNKLNSKNLLPAYRNPSPRA 543
QY 362 --SLEKPGGHNKCFKACVAGSALQIHLNHTGERPYKCVNGCR--FTTGNIKLVH 417
Db 544 EAGLAPRPTGYRCLGCDATSLKSLARHYDRSRRIEYTCNCAIRLYFPNCSLLHA 603
QY 418 HHRERK-----YPHVQNNPVRPHLDVITSSGRLPYGMSVPEKAEEAGTGGVERK 472
Db 604 REHKDKGLVWCOSHLVNRVALDQWVGORDITPLRP--VAVPR--VSGPLALPALGEG 659
QY 473 PLVASTALSATESLTL-LSTGTSTAVAPGLPTENKEVYLMKAVERPSKADENTPPSGGS 531
Db 660 AITSSAITTVAAEAPVPLST-----EPRAAPATSAUTCFCLECKECCGDK-----706
QY 532 ATAGVA-----DSSATPMOLSKLVTSLPsmallTLNLKSTGSPFPVYLEPLGASP 583
Db 707 --AGMAAHFOQLGPRPAGATSNVCPCTPMILPNKCSFSNHRNKNRP-PHVCPECGG--761
QY 584 SETSKLOOLVEKIDROGNAVAVASTAGAPTTNAPAPSSASAGNOCVICTRLVSCPRALR 643
Db 762 ---NFLQANQOTHLREACLNHSRVG-----YKRPCSGVVGCVNSIK 801
QY 644 LHYGNGGERPCKVCGRAFSTRGNLRAHFVGHKTSAPAARONSCPTCOKKFTN 698
Db 802 SHIOTSHCEVFNKCRPTPMAFKSGPSAHNL--YSQHPEFQTOAKLILYKCAMCDTFTTH 859
QY 699 AVTLQOHVNMHLGQIPIPGSALSSEGGAAOENSSBOSTAGSGSPPOSOQSPREEM 758
Db 860 KPLLSHFDOHL--LP-----ORVSYFKRPSCLLPAQKRTML-----EHL 898
QY 759 SEEEDEEEEDVTDSDSLAGRSESGEKAISVNGDEEVSGAEDEVATSVAPTTVK 818
Db 899 KNTHOSGRLE-----TAGKGA---GALLLPKTEPEELA-----VSGGGAAPATEE 942
QY 819 EMDNSKAPQHTLPPPPPPDNLHPQMEOGTSDVSGAMEEAKLEGISPMALALQEG 878
Db 943 SSSSSS-----EEEVSSPSPRPAKRRP--RELGSKGLKGG-----GGG 980
QY 879 EGTSTPLVEELNLPKAMKKDRGESSGRACVCGSPPTQATALEHOK--TRPKD 931
Db 981 PGGWTCGLCHSWPBERDEYVAHMKREHGKSVKKFPCRLCERSCSAPSLRRHVVNHNIGI 1040

AC Q9H8L4: 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
ID 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
CDNA FL13479 f1s, clone PLACE1003738, weakly similar to zinc finger protein 135.
DE
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN
RP
PP
RS
RT
CC
EMBL: AK023541; BAB14602.1; .
HSSP: P25490; 10BD.
InterPro: IPR000822; Znf_C2H2.
Pfam: PF00096; zf-C2H2; 13.
PRINTS: PR00048; ZINC_FINGER.
ProDom: PD000003; Znf_C2H2; 2.
SMART: SM00355; Znf_C2H2; 13.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 13.
PROSITE: PS50157; ZINC_FINGER_C2H2_2; 13.
DNA-binding; Nuclear protein; zinc-finger.
SEQUENCE 498 AA; 54872 MW; C704FCF55CCDAELA CRC64;
SQ
Query Match 6.1%; Score 321.5; DB 4; Length 498;
Best Local Similarity 22.1%; Pred. No. 2,4e-11;
Matches 141; Conservative 70; Mismatches 237; Indels 191; Gaps 24;
371 KCRCAKGVESALQIHLRSHTEGPERPCVCGNRPFTTGNLKVHFRHREKPHYQWN 430
20 KCAKCPKAYGALSKIKIHQKHTGERPRACADCGKSPADS---VFKKHR--THAGIR 73
431 PNPVPHLDVYITSSGLRPGVMSVPRKAEAEAGTPGGVGERKPLVAST--TALSAETSLT 488
74 P-----YSGERGKAAVA-ELKDLRNHERSHT-----GERPFLSCSGKSFSSSSSLT 119
489 LLSTGTSAAVAPGLTFPNK--FVLMKAVPERKSKADENTPPSGSGAIVAGVDSGATRMQL 547
120 CHORIHAAQKPYRCACGKGTOLSSYOSHERTHSGEKPLCPRCGRMFSPSSFRHQR 179
548 SKLVTSLSWMLLTNHLKSTGSPRPYVLEPLG---ASPSKTLQOLVERKIDRGAVAV 604
180 AH-----BGVK-----PYHCEKCGKDFROPADLA-MHRRVHTGDR----- 213
605 ASTASGAPTSAPAPSSSASGNQCVITCLVLSCPRALRLHYGNGHGERPFKCKVCGRAF 664
214 -----PRKCLQCKTFFVASMWLKRALVHSGRPFCEGEGRAF 252
665 STRGNLRHAFVGHKTSPPARAQNSCPLCKKFTNAVTLQGVNHLGQ---IPNG---- 717
253 AERASLTGHSRVH---SGERPFHCNACGKSFVSSSLRKHERHRSSEAAGVPPAQLV 308
718 -GSALSEGGAQOENSSQSTASGSPFQPOSGQSPPEEMSEDEDEDEEDVTD 776
309 VGLALPV--GVAGESSAAAPAGAGLDPPA----- 336
777 SLAGSGSSEGEKAIIVRDSEFVSGAESEAVATVAATTVKEMDSNEKAPQHTLPPRP 836
337 GLLGLRPESGG-----VMATQGVVGM----- 358
837 PPDNIDHPQPMQGSIDVSGAMEEELKLEGISPMALITQGEQSTPLVEELNLPKAK 896
359 ---TVEHNEJECODAGVAREAGPL-----BGAG-----EAGG 385

OY 897 KDPGESSGRKACEVCGSPPTQTALEHOKTHPKDGLFTFCVFCRGLDRATLKKHML 956
DT
ID
AC Q9NS43: PRELIMINARY; PRT; 751 AA.
ID Q9NS43
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical zinc finger-like protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN
RP
PP
RS
RT
CC
EMBL: AC074331; AAF88107.1; .
HSSP: P08045; 1ZNF.
InterPro: IPR001909; KRAB.
Pfam: PF01352; KRAB; 1.
PRINTS: PR00048; ZINC_FINGER.
SMART: SM00349; KRAB; 1.
PROSITE: PS0805; KRAB; 1.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 18.
PROSITE: PS50157; ZINC_FINGER_C2H2_2; 17.
DNA-binding; Metal-binding; Nuclear protein; zinc-finger.
NON-TER 1
SEQUENCE 751 AA; 86207 MW; 96583BA581C87FC CRC64;
SQ
Query Match 6.0%; Score 320.5; DB 4; Length 751;
Best Local Similarity 20.3%; Pred. No. 4,6e-11;
Matches 158; Conservative 71; Mismatches 287; Indels 263; Gaps 23;
231 SSTKPLPLFSPRIKPAOTGKTTASSSSSSSSGAEPKQAFENLYNPLGSOHPFVGVG 290
171 TDEP-----KPKGMEYGIITISDSNOKLPLGEK-----HPCGCGSG 210
291 RSHKTPAPSPALPGSDIOLASPHLAFPTGTGLLAOCLGARGLEAASBGLPK-- 348
211 PSY-----SPRLPLPN--VHTGEKCFQSGSHLRTQRIHNGEKILNRCHESGDCEFNKS 262
349 -----NGSGELTGYGEVYISLEKPGKHKCFCAKVPSSDSALQIHLRSHTEGPERPKCNV 402
263 FHSYOSNHTGEKSY-----RCDSCGKGFSSSTGLIHYRHTGEKPKCEE 308
OY 403 CGNRPFTGNLKVHFRHREKPHYQWNPHVPHRELDVITSSGLRPGVMSV----- 453
DB 309 CCKCFSSQSNFQCHQVHTEK-----YKCECGGFGFSGVNLRYHQRVH 354
OY 454 ---PREKAEAEAGTPGGVGERKPLVASTALSAATESLTLSTGTSAAVAPGLTFNKFVL 510
DB 355 RGEKPYKCEE-----CGKGFQAANFHT 377
OY 511 KNAV---EPKSKADENTPPSGSGAIVAGVDSGATRMQLSKLVTSLSWMLLTNHLKST 567
DB 378 HORVHTGEKPYKCD-----YCGKGFSHNS-----PLCHRRVHT 411

[illegible][illegible]

01-JUN-2002 (Tremblrel. 21, last annotation update)
Unknown (protein for MGCL12654).
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
Tissue-uterus;
Strasbourg R;
Submitted MAY-2001 to the EMBL/GenBank/DBJ databases.
EMBL; BC007777; AA07777.1;
InterPro: IPR003309; Tr9-SCAN.
InterPro: IPR000822; Znf_C2H2.
Pfam: PF02023; SCAN; 1.
Pfam: PF00096; Zf-C2H2; 13.
ProDom: PD00003; Znf_C2H2; 6.
ProSITE: PS0804; SCAN_BOX; 1.
ProSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_13.
ProSITE: PS00157; ZINC_FINGER_C2H2_2; 13.
DNA-binding; zinc-finger.
SEQUENCE 734 Aa; 82055 Mw; B804766A0B008048 CRC64;
Query Match 5.98; Score 312; DB 4; Length 734;
Best Local Similarity 19.4%; Pred. No. 1.4e-10;
Matches 204; Conservative 75; Mismatches 329; Indels 444; Gaps 39;
11 LGPGCEPARERG-----DASEHHPOV-----CAK-----CAQFSD---PTEFLAHQ 51
6 LGSPDRAPPEDECPVAVKLEDSDEGEALMDPGEAPLARLRRCRYEATPQDEALQL 65
52 NSCCDDPYVVIIGQENPSSASSAPREGHSRSQVMDT-----EHSNPPD---SGSS 103
66 RELC-----RQWLREVRKSEQLLELLVLEQFLGALRPETIARYQ 105
104 GPPDPTWBERGEESGQFLVATGTAAGG-----GGLIASPKLGT---PLPP 152
106 G-----QRGSPREAAALVDGLRREPGRRRWVVOVGQEVLSKEMEPSSQPLPE 157
153 ESTAPPPPPPPPPPPVSGSHNIPILBELRYLQRO-----HQOMTQICQVLL 207
158 TERPPRPGKPP-----RTQESPLDLOKVESEVED----- 192
208 LGSIGQVGAAPSSELPETGTAASSTRKPLPLPSPIKAQTKTATSSSSSSSGAEP 267
193 -----SDFESSGLPAATQESVPTLLP--EAGRCGVLDQIFPHSKGPPEP 237
268 KQAPPHLHPGSGHPFSGVGRSHKPTPAPSPAL-----PGSTDQLIASPHLAP 319
238 SWR-----EHPRALMHE-EAGGIF-----SPGFALQGISAGPGS-----VSPHLHVP 280
320 GTTGL--LAACIGAGRGLEAAASPGLLKPKNGSGE-----LGYEVISLEKP 366
281 WDLGMAAGLSGQIQRSR--EGGFALHLLPSDLRSDEDPDEDPCRGVAPLITRWNSP 338
367 GGRH-----KCRCAKYFGSDSALQIHLRSHTERPYKCNVCNRPRTTRCN 412
339 RGRSGRGPSTGGGVVGRGCDVCGKVFQSRNLLRHQKHITGERPFVCSGGRSRSRH 398
413 LKVFHRRREKYRHYOMNHPVREHLDYITSSGLPYGMSVPRPKAEAGTGGVERK 472
399 LLRQLQHLTERPPV-----CGDCGGG--- 420
473 PLVASTALATESLTLSTGTSTAVAPGLPTEPKFVLMKAVEPKSADEMTPPSEGA 532
421 -----IVRSARLEHRRVHNGEOP----- 439
533 IAGVADGSATRMOLSKLVTSLPSMALLTNHLKSTGSPFPYVLEPLGASPSSTKLOQL 592
440 -FRCAECGQSPFRQSRN-----LLOHQRIHGDPGPR-----GAKP----- 472
593 VEKIDRGAVAVASTASGAPTTASAPSSA-----SGNQ--CVTCL 633

473 -----PAPGAPPPPPFPCCSECFESFARRAVLEHQAIVHTCDKSEFCYECG 519
QY 634 RVLSCPRALRHGQHGGERPFCKVCGRAFSTRGTLRAHPVGHKTSPPARAONSCPTCQ 693
Db 520 ERGGRSVLLQHRHVHSGERPFACGCGSPFRQSRNLQHRRIH-----TGERPFACGCG 575
QY 694 KKFTHNAVTLQOHVYMHILGQIPLNGSALSSEGGAAOENSSEOSTASGPGSPPOQOOPS 753
Db 576 KAFRRQPTLQHLRVHTG-----EKP- 596
QY 754 PEDEMSDEEEDDEEDVDTEDESLAGRSESGEKALISVRGDSERVSGAEEVATSVAA 813
Db 597 -----FAC 599
QY 814 PTVTKEDSNKAPQHTLPPPPPPNDIHPQMEQTSQSVSGAMEEAKLEGISSPMAA 873
Db 600 PEGGQFSSQRLKTLRHQRTHTGKPP---YHGEGCGLGFTQVS----- 638
QY 874 LQEGSGTSPVLEENLPEMKKDPGSSGRKACGCGSPFTQALAEHQTKHRKDP 933
Db 639 -----RLTEHQRIHTGERP--FACPEGGQSPRQHANLTOHRRITHGERP 680
QY 934 LEFTCFRQGFDRATLKKHMLAHQVPPA 965
Db 681 -YACPEGCGAFRQRPPLTQH-LRTHRRKRPRA 710

RESULT 32
Q9V009 PRELIMINARY; PRT; 744 AA.
Q9V009;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Bowl protein.
GN BOWL OR CG10021.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta.
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anagnostides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaziel R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abill J.F., Agbayani A., An H.-J., Andrews-Piannkock C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Burkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Switkes R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*,"
Science 287:2185-2195(2000).
EMBL: AE003578; AAF51065.1; -.
DR HSSP: P07248; ZADR.
DR FlyBase: FBgn0004893; bow1.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf_C2H2; 5.
DR ProDom: PD000003; Znf_C2H2; 2.
DR SMART: SM00355; Znf_C2H2; 5.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 5.
DR DNA-binding: Metal-binding; zinc-finger.
KW SEQUENCE 744 AA; 79831 MW; 403653CE7F57672D CRC64;

Query Match 5.9%; Score 311.5; DB 5; Length 744;
Best Local Similarity 18.1%; Pred. No. 1.5e-10;
Matches 193; Conservative 90; Mismatches 281; Indels 505; Gaps 37;

QY 90 MDTFHSNPDSGSSGPPPTWMPERGEESSGQFLVA-ATGTAGGGGGLILASPKLG-- 146
DB 1 MPTSSSEISGGGGAIPMLRPSR-----MDQFMNSMAAAAAAVGGGLPGAADNRNGSS 55
QY 147 -----ATPLPEESTP-----APPPPPP----- 164
DB 56 GGSDDGSGNGNDSSNSASRISAYETOLAYOQHLAGLHGRPPRRPHHRIISAFVPL 115
QY 165 ---PPPEGVGSHLIPILLELVLYOQROIHQMTDQICQVLLSLGQTVAPASP 221
DB 116 PTGVPRPSNSNYEITAMADKKELALRE-----AAAAAAMLGKPGCGP 163
QY 222 SELPDTGAASSKPLLPSPKPAOTGKTATASSSSSSSGAEPKKAFFLYHPILGSG 281
DB 164 G-VPEPPGV-----LTGP-----AGVPPP----- 180
QY 282 HPFSVGVGRSHKPTAPASPAALPGSTDIASPHLAPPTGTLAAOCL--GAAGGLENA 339
DB 181 -PYLTG-----PGSPPTGAGS-----PPFPGAAALAFPELGGGMHNG 219
QY 340 ASPGLLK-PRKNSGELGYEIVSSLEKPGGRHKCFKAKVFGSDALQIHLSHTGERPY 398
DB 220 LDRRLRAPRGRAS-----RPKKQIFCFCNKPFTKSYNLLIHRTHTDERPY 266
QY 399 KCVNCGNFTTNGNLKVHFRHREKYPHYGMNPRVPEHLDLYVITSSGLPYGMSVPEKA 458
DB 267 SCDDCGKAFRRDDHLMDH-----RYIH----- 288
QY 459 EEEAGTPGGGVERRKPLVASTALSTATESLTLLSTGTSTAVAPGLPTFNKFLMAKAVEPKS 518
DB 289 ----- 288
QY 519 KADENTPPGSESGAIAGVADSGSATRMQSLKLVTSLSFWALLTNHLKSTGSTPPRYLVE 578
DB 289 ----- 288
QY 579 LGASPSSTSKLQQLVEKIDRQGAVAVAASTASGAPTTSAAPASSSASGRNOCVILRLVSG 638
DB 289 -----SKKKPKCTECGKGFCQ 305
QY 639 PRALRLHYGQHGGERPCKKVCVGRASTRGNLRAHFVGHKTSPPARAONSCPIQCKKFTN 698
DB 306 SRTLAVNHLKILHDESPHKPCVCSRSFNGRSLKTHLLTN-----TDHKPEYCSGCAVFR 361
QY 699 AVTLOOHVAMHLGGQIPNGSA-----LSEGGG 726
DB 362 NCDLRRLHATLHAVGEVNSGDYVDVGEDEARNLSGDEEDSLLEVDSPROSPVHNILGESGG 421

QY 727 AAOENSSPOSTASGPGSFPOPOSQSPPEEEMSEEEDE-----EEEDVDYTD 774
DB 422 SGKESSEMERMLKRAALTHEES-----EEFDFDEEELDLPRLVINDLPREEDDDFP 476
QY 775 ED-----SLARGSES-----GGEK-AISVRDSEEV 800
DB 477 EDEQAEVALVARFQSKAAATSSSSSVGTKPRQGVTHCHNGEGEYTYMHPGEEKHQE 536
QY 801 SGAEDEVATSVAPPTVTKMDSNEKAPQHTLPP-----PPPNLDHP----- 844
DB 537 EPNGSIASLVPVPSPFVR-----YSVPGAAGPPRAPGAPRTHQHPHPLLP 586
QY 845 -----QPEEGGSDVSGAMEEBAKLEGI-----SSPMALPQEGESTPLVE 887
DB 587 PNGDPYLPILHYRDLHNSLSLNSKA-----GVPPRPHTPRTIITQPEEG--KPPNQ 636
QY 888 ELNLP-EMAKKDPGESSGRACVCGSQSFTQTALEEHOKTHPKDGLFTCVGSGQFLD 946
DB 637 PLHSPHEAMPFSLGSIPMKRI-----LPAPTLDIMDPHN--HPGLG-----QRTFVD 682
QY 947 RATLKKHMLAHQVPPFAPHG-PQNIATLSLVPGCSSISFSGLSPP 994
DB 683 SPSTVALNMSRH---PRLDLGKPPSTETSGATTKEGPPVAPPIAPP 728

RESULT 33
046205
ID 046205 PRELIMINARY; PRT: 1920 AA.
AC 046205;
DT 01-JUN-1998 (TREMBLrel, 06, Created)
DT 01-JUN-1998 (TREMBLrel, 06, Last sequence update)
DT 01-MAR-2002 (TREMBLrel, 20, Last annotation update)
DE zinc-finger nuclear protein HINDSIGHT
GN PEB OR HINDSIGHT OR EG:66A1.1 OR CG12212.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephyrgota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyrgotidae; Drosophilidae; *Drosophila*.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97330681; PubMed=9187140;
RA Yip M.L.R., Lamka M.L., Lipsitz H.D.;
RT "Control of germ-band retraction in *Drosophila* by the zinc-finger
protein HINDSIGHT,"
RL Development 124:2129-2141(1997).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: U86010; AAB95640.1; -.
DR HSSP: P07248; ZADR.
DR FlyBase: FBgn0003053; peb.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf_C2H2; 12.
DR PRINTS: PR00048; ZINC_FINGER.
DR SMART: SM00355; Znf_C2H2; 13.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 11.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 10.
KW DNA-binding; Metal-binding; Nuclear protein; zinc-finger.
SO SEQUENCE 1920 AA; 208973 MW; BA51E67584916AA2 CRC64.

Query Match 5.9%; Score 311.5; DB 5; Length 1920;
Best Local Similarity 18.8%; Pred. No. 4.9e-10;
Matches 259; Conservative 134; Mismatches 460; Indels 527; Gaps 54;

QY 6 GSSSRLLGPGCGEPAERKGDSEHNPOVCACACAOFSPTERFLAH-----QNSC 54
DB 226 GSPQ--GQCLSSGESGIGADEHMKYLCPICEVVSATPRHEFTNHIRCHNANGTENT 283
QY 55 C-----TDPRVNTIIG----- 65
DB 284 CRICKSVLSSASLDRHVLVHTGERPFCRCYCHLTFTTNGNMHNRHTKQHOVAOSQ 343
QY 66 -----GQENPS-----NSSASSAPRP 81

344 SQQQOOSLQOOOOSQOOORQOOQHOPOSOQOOQNPAAQOOLMGNLTLSAGAESYEDSASCSTDVS 403
EGHSQVMDTEHSNPDSDGS-----SGPPDPTWGPERR----- 115
404 SGHSRSHRSSSSLLNNNNNNNNKANNLNKDLLEELEVSTEDODTENKORLKTITNNNITIESE 463
116 -----GEESGOFVAATGTAAAGCGGLIILSPKLGATPLPESTPAAPPPPPPPPP-- 167
464 QOEDMDDEADADVAMLTJTP--DVATLLAGASASGA-----AASRTTPSPASAPLL 515
168 ---PGVSG-----HNTPLILEELRVL-----OQROIHQOMQTEOICR-Q 204
516 LSCPACGASDFETLPALCVHIDMAMHSDIPACRCDEVIJFATHQLOSH-----CCRLP 568
205 VLLGSLIGOTVGAPASP--SELPGTAASSTKPLPLPSPIKPAQOGKTTASSSSSSSS 262
569 NALAGGLPPLGLASSSPDLNHEEPDEHGDDEDL-----EOKERLASOSED-- 614
263 GAEPKQAFPHLYHPLGSOHPFVSGVGRSHKPTP-----APSP 301
615 -----FFHOYLKHKTKTANGCGAISHPPSPDKHEPADTKDLADIOSILNMTSSSS 663
302 ALPGSTDOLIASHP---LAPPTTGLLAAOCLGAA-----RG-----LEAASPGLLK 346
664 SELRNFEOSVNTPNSSOYSLDGRDOEEBAQDAFTSEFRMKLRGEPPCKLCTAVFPLRA 723
347 PKNGSGELGYEYISSLEKPGGRHKRCRCARVFGSDSALOIHLSRSHGERPYCNVCNGNR 406
724 LK-----GHNRVHLGAVGAPAGAPRCMCPYAVCDKALVHRHRTINGDPRPECAVCNYYA 777
407 FTTGRNLKVFH-HRH---REK-----YHVOVMP 431
778 FTTKANCERHLNRHGKTSREEVKRAIYVHPADAGCEDSKRGLGDLADTSFRSISPTP 837
432 HPVP-----EHLDYVTSS--GLPYGMSVPEPKAE 459
838 PPPPVNESKSQILKMLLGENHLPVNOORPLKIOYKSLDQVLDKRPAPRQOOQOOQOO 897
460 EEAQT-----PGGVERKPLVASTALSTESTLSTGSTTAVA----- 499
898 EKSGSALDESMVDLSTKRTGASLTPAVYRTTPPAVAAPVTPGCGVGTBDLAAITBOO 957
500 -----PGLPTFNKPYLMKAVEPKSKADENTP----- 525
958 LLLAQQOLFGAGGEYMQOLFERSIMFQSGTSGFP---FFPFMAAPPPQANPEKPPMWSPP 1013
526 ---PGSEGSALTAGVADSGATRMQLSKLVTSLSMALLTNLKSTGSEFPFVYLEPLGA 581
1014 NRINMPYGVGVGPVPPGQPVKVIKNGV-----LMPKQKORRYRTERPPACERCSA 1066
582 SPSESTKLOLVEKIDRQ-----GAVAVASTAGAPTTAPAPS 620
1067 RFTLRSNNERNHKOQHPOFYAQRERSAHNHWRGASVNAALAAAAAARVYVMAGGPG 1126
621 SSASGPNOCVICLVLSLSPRALRYHGO-HG--GERP---KCKVCGAFTSTRGNLRAH 673
1127 SSGGGSNNH-----HHGHGSHGHAPISEQVKAIIAQ-----QKKAH 1164
674 FVGHKTPAARONSCRPICOKFTNNAVTLQOHNVMHLGGQIPN-----GGSALSEGGA 728
1165 -----KNTDILLOALAHGSSSVAGNPLLHFYPLTNPSPIHNGS--SOGGOA 1210
729 -----QENSSEOST-ASGPGSPFPOSOQSPPEEKSEEBE-----DEE 768
1211 TAMDDDEKLITDEDENEHDHVEAEVDDF-----EEDDEBEEDDEPEDEPLILDEOP 1265
769 EEDVTDSDS---LAGRSESSEGE-----KAISVGRDSEVVSAGEEVEAT 809
1266 AEKEAEHEQLEPKPLFOLGTKEAAQKMAETILFOAIKAGKPLSPPTKEMASPANPTVAT 1325
810 ---SVAPPT-----VKEMDSNE-KAFQHTLPPPPPPDNDHPQ 845

Db 1326 TMOEPAITAPSTNPSSLKTIWIAOAEYVCKSLKEVASSPEFKDESODLYPAKLVDNAT-SQ 1384
QY 846 PMEGCT----SDVSGAMEEAKLEGISSPMNALQEGEGTSTPIVIBELNLPKAMKNDPGE 901
Db 1385 NMGFNSYRPPSDVANAHEOSDE--EGLVAGSGASSENNGSTEDVTSSSSSEPKKSAVSL 1443
QY 902 SSGRACCEVCGQSEFPOTALHEHQKTHPKDGPLFTVCVCRGOFPLDRATLKKHMLAHQV 961
Db 1444 APNRVSCPYCQMFPPWSSSLRRHILITHTGOKP-FKCSHCPLPLFTTKSNDRHILRRHGNV 1502

RESULT 34
061360
ID 061360 PRELIMINARY; PRT; 962 AA.
AC 061360;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CROL alpha.
GN CROL OR CROL ALPHA OR CG14938.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRATIN=CANTON S;
RA D'Avino P.P., Thummel C.S.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CB EMBL; AF020347; AAC15516.1;
DR HSSP; P07248; IARD.
DR FLYbase; FBgn0020309; crol.
DR Interpro; IPR000822; znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 18.
DR PRINTS; PR00048; ZINC_FINGER.
DR ProDom; PD000003; znf_C2H2; 5.
DR SMART; SM00355; znf_C2H2; 18.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 18.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 18.
KW DNA-binding; Metal-binding; Nuclear protein; zinc-finger.
SQ SEQUENCE 962 AA; 106360 MW; 618D1BC9799C9AAB CRC64;

Query Match 5.8%; Score 310; DB 5; Length 962;
Best local Similarity 20.1%; Pred. No. 2; 6-10;
Matches 177; Conservative 83; Mismatches 321; Indels 300; Gaps 33;

QY 161 PPPPPPPGVSGGHLNPLLEELRVLQORQIHQMOMTEQICROYLLGSLGQTGVA--- 217
Db 89 PPPPLLP-----QVTSNAAASAAAAAATNNNAAVA 121
QY 218 -PAPSELPGTCAASTKRLPLFSPIKPAQTG---KTIASSSSSSSSSGAB----- 265
Db 122 VMASANAAAAAASAGGGLP-----PATSGNGGOQVTVTTSSSTSSGSGTSGTT 175
QY 266 -----PRKQAFPHLYHPLGSOHPFVSGVGRSHKPTPAP--SPALPGS----- 306
Db 176 TTAGELMPKMEGHIHGVDSGN-----GGNGGQVNALAPDGTPIATGTIVCDICGM 229
QY 307 ---TDOLIASPHLAPPGTGLLAOCLGAARGLEMAASPGLLKPRKNGSGELGYEYI--- 360
Db 230 FQFRYQLIV--HRRYHSEKRFWCQVCG--QGFTTSD--LTRHGKIHIS-GPWFCTI 280
QY 361 -----SLEKPGGRK-----CRCAKYVGSALDILHLSRSHGERYKVCNCGN 405
Db 281 VCFNVFANNSTLERMKRKHSTDKPFACCTICQKTFARKENHLDNHFSHGTEPFRQOYCAK 340
QY 406 RFTTRGNLKVHFHRRREKYPH---YQMNPAVPEHL--DYVITSSGLPYGMSVPEPKAE 460
Db 341 TPTKKEHVNVHRRKTTGETPHKCDICKSFTKKEHYVNHVYHNGQTTHOCVCGKK--- 397
QY 461 EAGTPGGVERKPLVASTALATESLTLSTGTSTAVAPGLPTFNKPYLMKAVEPKSKA 520

Db 398 -----YTRKEHLAN-----HMRSHT 412
QY 521 DENTPPSSGSAIAGVADSGATRMQLSKLVTLSPSWALLTNHLKSTGSPRPVYLEPG 580
Db 413 NE-TPFRCE--ICKSPSRKEH-----FTNH-----ILMHTG 441
QY 581 ASP-----SEFT-SKLOOLVEKIDROGAVAVASTASGAPITTSAPASSSAGSPNCVCIL 633
Db 442 ETPHRCDFCSKTFTRKEHLNHNV-RO-----HTGSPHRCSTCM 479
QY 634 RVLSCPRALRLHYGONGBERPFCVKVCGRAFTSGNLRAHFVGHKTSPARAONSCPTCQ 693
Db 480 KTFPRKEHLNVNHIHQHTGETRPFCYCTKAFTRKD-----HMVNVHROHTGSPHKTCTCT 535
QY 694 KKFTNAYVTLQOHVNRHNLGQOITNGGSALESGGAOENSSSQSTASGFCSPPOQSOQPS 753
Db 536 KTFPRKEHLNHNVROHTG-----DSPHRCSTYCKKTFTRKEHLTNH 575
QY 754 PEELMSSEEEDEDEEDVTDDESLAGSGESGGEKALSVRGDSEEVGAEVATSVAA 813
Db 576 VRLTGDSPHKECYCKTKTTRKEHL-----NNHMRQSSDNPHCCNVCK 620
QY 814 PTVKEMDSNEKAPQHTLPPRPRLDHPQPMEOGTSDVSGAMEEBAKLEGISSPMAA 873
Db 621 PFTREHLINHMRSCHTGRDP----- 641
QY 874 LTQSGEGSTPLVEELNPEAMKKDRGESSGRKACVCGSFPTOTALDEHOKTHPKDGP 933
Db 642 FTCTCGSFPRLKGNLFHQRSHTKQOEMERPFACERPKNFICKGHLVSHMRSGEKP 701
QY 934 LFTGVCROGFLDRATLTKKMLLHNHVV-----PPFAPHPQ 970
Db 702 -HACTLCSKAFVERGNLKRHKMKNHNDAMPPRPVH-PO 740

RESULT 35
Q960L6 PRELIMINARY: PRT: 736 AA.
AC Q960L6: 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE LD46233P.
CG122299.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_Taxid=7227;
OX R N
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunco J., Paciel J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.,
RA Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY051994; AAK93418.1;
DR FLYBase: FBgn0032295; CG12299.
DR InterPro: IPR000822; znf.C2h2.
DR Pfam: PF00096; zf-C2H2; 10.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_10.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 10.
DR DNA-binding; Zinc-finger.
SQ SEQUENCE 736 AA; 81870 MW; 0B1C198D547B4C2C CRC64;

Query Match 5.8%; Score 309; DB 15; Length 736;
Best Local Similarity 20.8%; Pred. No. 2.1e-10;
Matches 174; Conservative 91; Mismatches 283; Indels 288; Gaps 33;
QY 142 SPKLGATPLPPESTPAAPPPPPPPGVSGHLN-----PLIL-BELRVLQQRQIH 193

Db 17 TPILQOHTLLPQARPNPQOOPQORPDLTFHMCSCAEFFVHPLATLYQHMTLPHNEBG 76
QY 194 QMO-----MTEQICROVLLIGSLQITGAPARSPSELPTGSAASSKRPILP 238
Db 77 NGQOQESPGESEEDYSWFEPYC-----ELAEDSDSSDGASGSDSSSSSD--- 126
QY 239 LFSPKIPAQGTKTASSSSSSSSSSGAEPPKQAFPHLVNPLISQHPFS--VGVGCRSH--- 293
Db 127 --DDDDDDDDSSSSSSSSSSSSSVPTTSNSNTQOQSQSVORPHLVLVAGPYNEPL 184
QY 294 -----KTPAPSPALPGSTDLIASPHLAFPTTGILLAAQCLGAAGLEA 339
Db 185 QMTDRESTSIFMVQTVSVTP-----LQOLLPRAPTVSPGLG--LQSTPIKRRGRASN 237
QY 340 ASPLGLKPK-NG-----SGELGGEVLSLEKPGGRKRCRCACAVFGSD 382
Db 238 IGAPVMDPALNGOKCFQOCTHCEASFPNAGDLKHNVRSHITNKP--FQCSICQKTFTHI 294
QY 383 SALQIHLRSHGTGRPYKCNVCGNRFTTGNLKVHFNHREKYPH--YQMPHVPVHLDY 440
Db 295 GSLNTHIRIHSGEKPYKCELCPKAFTQSSSLVMHMRSHSVKRPQCVCQCDKG----- 346
QY 441 VITSSGLPYGMSVPREKAEBAAGTPGGVERKPLVASTTALSAVESLTLLSTGTAVAR 500
Db 347 FINYSLS-----LLHQKTHIAPTE-----TFICP 370
QY 501 GLPTKFKFLMKAVEPKSKADENTPPSGSALAGVADSGATRMQLSKLY-----TS 553
Db 371 ECE-----REFKAEALLDH-----MRNHTQELVYQCAICREA 403
QY 554 LPSWALLTNHLKS-TGSEPPRYVLEPLGASPSETSILQOLVEKIDROGAVAASTASGAP 612
Db 404 FPASSHLVQHMKNHMEKPTCSL-----CDR-----SFTQSG-- 436
QY 613 TTSAPAPSSSASGPNOCVICLRYLSCPRALRLHYGONGBERPFCVKVCGRAFTSGNLRA 672
Db 437 SLNIHMRHITGEPKFOCKLDCKTCFQASLSLVHMKIHAGEKPYPCPICGSYQOAYLWK 496
QY 673 HFVGKHTSPAARQNS-----CPICQKRTNANTLQOHN-ML----- 710
Db 497 H1QAHOMASAAASSTPSGLLVAKQPHETLYCTGCSLHADATALLASHVSHQNAALLDTMK 556
QY 711 -----GQOIPNGGSALESGG-----AAOENSSSQSTASGSPSPPOQSOQPSPE 755
Db 557 QSGMNTAPGCAITDYKCSABEQAYVERVQCVLQAMNHOQOHOQOQPPQOQOQHPQOQ 616
QY 756 EEM-----SEEEE-----DEEDED-----V 772
Db 617 QOOHLLQOQPHQMLPQPKLPAMDSTGEDEBERDEAEPRDEEEDQDEBERPAEVKTEVL 676
QY 773 TDEDSLARGSGESG--EKAISVRGD-----SEVSG--AEEVAT 809
Db 677 AAEDALTNPGYPIILGIEQIILDSDMYEDFGDMVCGCEVFGGVVNEEEVYT 732

RESULT 36
Q9VKF1 PRELIMINARY: PRT: 934 AA.
AC Q9VKF1: 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CROL protein
GN CROL OR CG14938.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_Taxid=7227;
OX R N
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;

```

MEDLINE=20196006; PubMed=10731132;
Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers J.H.C., Blazer V., Champagne M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abbil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey K.A., Helman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Klimel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Meltz B., McIntosh T.C., McLeod M.P., McPherson D.,
Mekulov G., Milshina N.V., Mobarry C., Morris J., Mostrel A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
Palazzo K.M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Soler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Syrkas R., Tecce C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
"the genome sequence of Drosophila melanogaster."
Science 287:2185-2195(2000).
-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
EMBL: AE003633; AAF53121.1; -.
HSSP: P07248; IARD.
FlyBase: FBgn0020309; crol.
InterPro: IPR000822; Znf_C2H2.
Pfam: PF00096; Zf_C2H2; 17.
PRINTS: PR00048; ZINC_FINGER.
PRODOM: PD000003; Znf_C2H2; 4.
SMART: SM00355; Znf_C2H2; 17.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 17.
PROSITE: PS00157; ZINC_FINGER_C2H2_2; 17.
DNA-binding: Metal-binding; Nuclear protein; Zinc-finger.
SEQUENCE 934 AA: 103010 MW: 814443847589B8 CRC64;
Query Match 5.8%; Score 309; DB 5; Length 934;
Best Local Similarity 19.9%; Pred. No. 2,9e-10;
Matches 175; Conservative 80; Mismatches 299; Indels 328; Gaps 33;
89 PPPPLPLP-----QVTSAAASAAAAAASNNAAVA 121
218 -PASPSELPGTGAASSTRKLLPLSPKPAQTG---KTTASSSSSSSSSGAE----- 265
122 VMASANAAMAAAAAASAGGLP-----PATSGNGGQOVYTTTSSSTSSGSGTSGT 175
266 -----PRKQAFHLYLPLGSHPFPSVGVGRSHKPTPAV--SPALPSS----- 306
176 TTAGELMKPKHEGGHGVGSGN-----GGNGGGQVVALPDGTPATGTNYCDICGM 229
307 ---TDOLIASPHLAPPGTTGLLAQICGAARGLERMAASPGILKPKNGSGELGYGEVI--- 360
230 FQFRYQLIV--HRRVHSEKRPFCVCG--QGFTTSD---LTHHGKTHIG--GPMFTCI 280

```

```

QY 361 -----SLEKPGGRHK-----CRCAKVFSGSDALQIHLRSHTEGPRKCNVCGN 405
Db 281 VCFNFANNTSLERNAKRRSHTDPRFACTICOTKPAKKEHLDNHFNHSHTEPRCOYCAK 340
QY 406 RFTTRGNLKVHFNHREKRYH---VOMNHPVPEHL--DYVITSSQLPYGMSVPERKAE 460
Db 341 TFTTRHEHVNHRKHNGEPFHRDCICKSFTRKEHYVNHMHMTGQTPHOCVCGKK--- 397
QY 461 EAGTGGGGERPLVASTALSLATESLTSLTGSTAVAPGLPTPKFVLMKAVEPKSKA 520
Db 398 -----YTRKEHLN-----HMKSH 412
QY 521 DENTPPSGSAIAGVADSGSATRMQLSVTSLPSWALLTNHLKSGSPFPYVLEPLG 580
Db 413 NE-TPFRCE---ICGSFSKHE-----FTNH-----ILMHTG 441
QY 581 ASP-----SET-SKIQQLVEKIDROGAVAVASTAGAPTTSAPAPSSASGPNOCVIT 633
Db 442 ETPHRDCSKFTTRKEHLNHY-RQ-----HTGESPHRCSYCM 479
QY 634 RVLSGPRALRLHYGGGGERPFCKVCGRAPSTRGNLRAHFGHRTSPARAQNSGPICQ 693
Db 480 KTFTRKEHLVNHITROHTGETPEFKCTYCTKAFTRKD---HNVNHRQHTGESPHKCTYCT 535
QY 694 KFTNAVTLQOHVRMLGQIPINGSALSEGGAOENSSEOSTASGPGSPQPOSQPS 753
Db 536 KTFTRKEHLNHYRLHTG----- 553
QY 754 PEEESSEEEDEEEDYTDDESLAGRSGSESGEKAISVGCDSSEVSGAEEVATSVAA 813
Db 554 -----DSPKCEYQCKTTRKEHL-----NNHMRQSSDNPHCCVNCN 592
QY 814 PTYVKEMSNEKAPQHTLPPPPPPNDLHPQMEQGTSDVSGAMEEAKLEGISSPMAA 873
Db 593 PFTTRKEHLNHSRCHTGPR----- 613
QY 874 LTQEGESTPLVEELNLPDAMKKDPGESSGKACVCGSGFPPTQALAEHOKTHPKDGP 933
Db 614 FTCEGCKSFPLKGNLFFQRSHTKGQEMERPFACPKFNICKGLVSHMRSHGKER 673
QY 934 LFTGVCFGROGFLDRATLKKHMLAHQV---PPEAFGRQ 970
Db 674 -HACTLCKAFVERGNLKRHKMHHPDAMPPEVHP-PQ 712
RESULT 37
Q9W4J1 PRELIMINARY; PRT: 1893 AA.
AC Q9W4J1;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, last sequence update)
DT 01-MAR-2002 (TRENBLREL. 20, last annotation update)
DE PEB protein (GHI0905P).
GN PEB OR EG:66A1.1 OR CG12212.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.H.C., Blazer V., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abbil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,

```


Db 525 CPQCGKRRVRRASLRLHLTH -GGRPRHNCTQCGSKSGFGOTDLRHNQSHHG----- 575
 QY 749 SQQSPPEEMSEEEDEDEDEEDVDDEDSDLAGSGSESGEKAISVRGSDSEVSAGAEVEVA 808
 Db 576 -EKPC-----RCSECG-----EGTSGASHLRHQ-- 598
 QY 809 TSVAAPTTVKEMDSNEKARQHTLPPPPPPPPDLHPQMEQIGSDVSGAMEBEAKLEGIS 868
 Db 599 -----RIHTGK-----PHACDTGCH-----RFRNS 620
 QY 869 SPMAALTOEGEGSTPLVEELNLPEAKMKDPPGSSSGKACEVCGSPPTQALAEHQKH 928
 Db 621 -----NLARRRSHGTGERP--YSCQTCGRSFRRNALRLRLHATH 657
 QY 929 PKDG-----PLFTVCFCROGFTDRATLKKNMLLHNHQVPPFA 965
 Db 658 AEPEQEOAEPPQEOECGSKFSRSCNLRLH-LLVHTGARPPYS 698
 RESULT 40
 Q9NT61 PRELIMINARY; PRT; 644 AA.
 ID Q9NT61; AC Q9NT61;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 71.4 kDa protein (Fragment).
 GN DKF2P564M2178.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP TISSUE=FROM N.A.
 RC TISSUE=BRAIN.
 RA Ottenwälder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL137516; CAB70782.1; -.
 DR HSSP; P07248; IARD.
 DR InterPro; IPR000822; ZnF_C2H2.
 DR Pfam; PF00096; zf-C2H2_15.
 DR PRINTS; PRO1600; ZNOCCLUDNS3.
 DR PRODOM; PD000003; ZnF_C2H2_1.
 DR SMART; SM00355; ZnF_C2H2_15.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 15.
 DR KW DNA-binding: Hypothetical protein; Metal-binding; Zinc-finger.
 FT NON_TER 1
 FT 1
 SQ SEQUENCE 644 AA; 71351 MW; FEA87E67E05049B1 CRC64;
 Query Match 5.7%; Score 300.5; DB 4; Length 644;
 Best Local Similarity 22.2%; Pred. 0.5; 8e-10;
 Matches 143; Conservative 75; Mismatches 246; Indels 181; Gaps 25;
 QY 193 HOMQMTEQ-----ICQVL-LIGSLGCTVGAPASPSSEL--PETGAASSTKPLPL 239
 Db 70 HQLOQLHRSREGVKPCPLCSRVRRPSSLDHLDHDSSEHFLCYDCCGALFTGTEALLLA 129
 QY 240 F-----SPKPAQTKGTATSSSSSSSSSSGAERPKQAFGLVLPRLCSQHFPFSGVGGRSH 293
 Db 130 HRRAHTPNPLHSCPCGKTFFVNLTK-----FLYH-----RKHGVGGVPLPT 170
 QY 294 KTPRAPSPALPGSTDQLIASPHLAPGTTGLIAAOLCGARGLEAA-ASPGLLKPKNKSG 352
 Db 171 TVPPREPP-----VIGPEPEPA-PAETGGEADEAPRVSEETSAPAAAGTVRCILCSR 221
 QY 353 ELG-----YGEVISLKEPGRBHKCFKAKVFGSDSLQIHLRSHGTGRRPKCNVCGN 405
 Db 222 EHGKALQTLRHQFVHRLER---RHKCSIGCKMKKKSHVNRHMLRHTGTGRRPPCPDCK 278
 QY 406 RFTTGNLKVHFRHREKYPH-----VQMPHPV--PEHIDVYITSSGL---- 447
 Db 279 PENSANLALRHRLTHHGGERPYRCGDCGKAFQSSSTLRHRLVHQAQVPRPCQEGVAFNR 338

